Aalto University
School of Science
Degree Programme in Engineering Physics and Mathematics

Toni Mäkelä

On variations of arclength with Myers's and Hawking's theorems in Riemannian and Lorentzian geometry

Special assignment in mathematics
Espoo, May 21, 2018

Supervisor: Dr. Kirsi Peltonen

Aalto University
School of Science
Degree Programme in Engineering Physics and Mathematics

ABSTRACT OF
SPECIAL ASSIGNMENT

Author:	Toni Mäkelä		
Title:	On variations of arclength with Myers's and Hawking's theorems in Riemannian and Lorentzian geometry		
Date:	May 21, 2018	**Pages:**	6+35
Major:	Engineering physics	**Code:**	MS-E1980
Supervisor:	Dr. Kirsi Peltonen		

We work through the first and second variations of arclength and discuss the rise of index forms and Jacobi fields along with their application in finding conjugate or focal points on Riemannian or Lorentzian manifolds. We then prove two theorems on the maximal distances of two conjugate or focal points along geodesics for manifolds that satisfy certain boundedness conditions for the Ricci tensor. These are Myers's theorem in Riemannian geometry and Hawking's theorem in Lorentzian geometry.

Keywords:	Differential geometry, variation, spacetime, Myers's theorem, Hawking's theorem
Language:	English

Aalto-Yliopisto
Perustieteiden korkeakoulu
Teknillisen fysiikan ja matematiikan koulutusohjelma

ERIKOISTYÖN TIIVISTELMÄ

Tekijä:	Toni Mäkelä		
Työn nimi: Kaarenpituuden variaatioista sekä Myersin ja Hawkingin teoreemista Riemannin ja Lorentzin geometriassa			
Päiväys:	Toukokuu 21, 2018	**Sivumäärä:**	6+35
Pääaine:	Teknillinen fysiikka	**Koodi:**	MS-E1980
Valvoja:	FT Kirsi Peltonen		

Johdamme kaarenpituuden ensimmäisen ja toisen variaatiolauseen. Käsittelemme indeksimuotoja ja Jacobin kenttiä, sekä näiden soveltamista konjugaatti- tai fokaalipisteiden etsintään Riemannin tai Lorentzin monistoilla. Todistamme kaksi teoreemaa, jotka liittyvät suurimpiin mahdollisiin geodeeseja pitkin mitattaviin etäisyyksiin kahden konjugaatti- tai fokaalipisteen välillä, kun Riccin tensori on rajoitettu tiettyjen ehtojen mukaisesti. Kyseiset teoreemat ovat Myersin teoreema Riemannin geometriassa ja Hawkingin teoreema Lorentzin geometriassa.

Asiasanat:	Differentiaaligeometria, variaatio, aika-avaruus, Myersin teoreema, Hawkingin teoreema
Kieli:	Englanti

"To confine our attention to terrestrial matters would be to limit the human spirit."

–Stephen Hawking

Acknowledgements

The author is grateful to Dr. Kirsi Peltonen of Aalto University for the supervision of this work as a special assignment in mathematics. Additionally, the author wishes to honor the work and memory of professor Stephen Hawking, the news of whose passing coincided with the writing of this text.

Contents

1. Introduction

The main topics of this text are Myers's theorem on Riemannian manifolds, Hawking's theorem on Lorentzian manifolds and the variations of arclength in a context covering both Riemannian and Lorentzian geometry. The two theorems are concerned with the maximal distance of two conjugate or focal points along a geodesic, given that the action of the Ricci tensor on a (timelike) vector is bounded below by a constant value. Since the current standard cosmological model includes the observation that our Universe is expanding, our physical motivation is in the general theory of relativity, where spacetime is modeled as a smooth 4-dimensional Lorentzian manifold[1]. The main application of Hawking's theorem is in proving the existence of a singularity on certain expanding or contracting spacetimes. This then restricts the maximal proper time that may be measured by any observer during the existence of a Universe satisfying the criteria of the theorem.

The text aims to bring closer together the viewpoints of geometry typically obtained on courses in both mathematics and relativistic physics. Modern courses on differential geometry will most probably discuss Riemannian geometry and leave pseudo-Riemannian and Lorentzian manifolds for independent study. In addition, Myers's theorem is chosen here as a topic which a student should well be able to comprehend after the completion of such courses. Correspondingly, a physicist who has studied general relativity might well have focused mostly on the local properties of Lorentzian manifolds. This might leave the study of global properties – which after all have to do with determining the structure of spacetime on the largest scales – for further studies on cosmology and mathematical physics. As a global theorem bearing resemblance to Myers's theorem, we discuss Hawking's theorem.

Since the nature of the text is as mentioned above, we have chosen a notation lying somewhere in between the usual conventions of mathematics and physics. Our notational choices have been made in pursuit of minimal ambiguity, such that the reader should be able to follow the mathematical nature of all objects in an equation with as little effort as possible.

We begin with a review of some central concepts in differential geometry in sections 2 – 6. This has been done equally for the sake of completeness as well as for the purposes of recalling some common definitions and introducing our notation. For a more detailed treatment, the reader is advised to consult for instance the books by Tu [1] and Lee [2]. These sections are followed by a brief summary of the basic notions of relativity and pseudo-Riemannian – in particular Lorentzian – geometry in sections 7 and 8. Some true-and-tested supplementary reading on these topics is provided for instance by Misner, Thorne and Wheeler [3], O'Neill [4] or the many other ones cited throughout the text. The reviews are done to the extent that is needed for the reader to get a grasp of what underlying machinery is needed for understanding the more detailed sections to come. The first and second variations of arclength along with the rise of Jacobi fields are worked through in sections 9 – 11, before turning into Myers's theorem in section 12. The rest of the text involves opening up Hawking's theorem in section 13 and discussing its physical implications in section 14.

[1]A Lorentzian manifold is a pseudo-Riemannian manifold with index one. This choice of index enables us to use one time coordinate along the spatial coordinates.

2. Manifolds

Manifolds are a generalization of geometrical objects to an arbitrary dimension $n \in \mathbb{N}$ with $\mathbb{N}$ being the set of all natural numbers, which includes all integers greater than or equal to zero. A *point* is a 0-manifold, *curves* are 1-manifolds, *surfaces* 2-manifolds and for higher dimensions we simply refer to n-manifolds. The manifolds considered in this text build upon the definition of topological manifolds stated below.

Definition 2.1. A **topological manifold** of dimension $n \in \mathbb{N}$ is a topological space that is *Hausdorff*, *second countable* and *locally Euclidean* of dimension n [1].

A space is Hausdorff if all distinguishable points $p, p' \in S, p \neq p'$ within it have separate *neighbourhoods* i.e. open sets that contain the respective points: $p \in U, p' \in V, U \cap V = \emptyset$, where $\emptyset$ is the empty set. Second countability means that it has a countable basis. For the space to be locally Euclidean of dimension n, we require that for all points with a neighbourhood U in the space there is a *coordinate map* (coordinate system) $\varphi : U \to \mathbb{R}^n$ that is a homeomorphism from U to an open subset in the real space $\mathbb{R}^n$ [1]. The combinations (U, φ) are known as *charts*. In addition to such maps, there are also other ones that are used e.g. for moving between two different manifolds or for their comparison. This is a topic we will return to in section 4, but for now we only need the definition of an *isomorphism* of two mathematical objects, including but not limited to manifolds.

Definition 2.2. Two objects A and B are **isomorphic** if there are the morphisms $f : A \to B$ and $g : B \to A$ so that we have the *identity maps* on A and B given by $f \circ g = I_B$ and $g \circ f = I_A$. The morphisms f and g satisfying this property are known as **isomorphisms** [1].

For the majority of this text we shall consider *smooth* manifolds. First we however need to define what it means for functions to be smooth.

Definition 2.3. Let U be a neighbourhood on $\mathbb{R}^n$ and $j, k \in \mathbb{N}$. A function $f : U \to \mathbb{R}$ is C^k at point $x \in U$ if

$$\frac{\partial^j f}{\partial x^{i_1} ... \partial x^{i_j}}$$

are continuous at x when $j \leq k$, $i_1 ... i_j \in \{1, ..., n\}$. A **smooth** or **infinitely differentiable** function is C^∞ [1].

Let U and V be two neighbourhoods on a topological manifold M so that $U \cap V \neq \emptyset$ is an open subset on M. Since M is locally Euclidean, we have the coordinate maps $\varphi : U \to \mathbb{R}^n$ and $\psi : V \to \mathbb{R}^n$ and hence also the two charts (U, φ) and (V, ψ). The charts are *compatible* if the two function compositions $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ are C^∞, when defined. That is to say, there exist smooth transitions between two coordinate systems for a point in a region where the two neighbourhoods overlap. Continuing on our cartography-inspired naming scheme, a collection of compatible charts $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is called an *atlas*. The index α is used here only as a simple reminder that the atlas consists of multiple charts, though in practice we cannot give a unique integer label to each element if a collection is uncountable. The atlas $\mathcal{A}$ is *maximal* if there is no larger atlas in which $\mathcal{A}$ is contained.

Definition 2.4. A **smooth manifold** (C^∞ manifold) is a *topological manifold* with a *maximal atlas* that is the differentiable structure on the manifold [1].

The definition of C^∞ functions given above is however restricted to those from $\mathbb{R}^n$ to $\mathbb{R}$. To extend our discussion to smooth functions with domains on general C^∞ manifolds, we give the following additional definition.

Definition 2.5. Let M be a smooth manifold of dimension n with a neighbourhood U containing a point x. A function $f : M \to \mathbb{R}$ is C^∞ at x if there is a chart (U, φ) about x and the function composition $f \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}$ is C^∞ at $\varphi(x)$ in the sense of definition 2.3. [1].

3. Vectors, covectors and tensors

Let M be a smooth manifold of dimension n. As it is locally Euclidean, we may always assign to each point $x \in M$ a Euclidean space $\mathbb{R}^n$ and represent the point x using its *local coordinates* $\{x^i\}$, $i \in \{1, .., n\}$. It is this feature that lets us introduce the tools of linear algebra into our calculations, even when considering manifolds that are not *globally* Euclidean. Hence, let us summarize some of the basics. We start by defining a *vector space* over an *algebraic field* $\mathbb{K}$.

Definition 3.1. A linear **vector space** V over an algebraic field $\mathbb{K}$ has the following properties for all $\boldsymbol{v}, \boldsymbol{w} \in V; a, b \in \mathbb{K}$.
(i) $\boldsymbol{v} + \boldsymbol{w} \in V$
(ii) $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$
(iii) $a\boldsymbol{v} \in V$, in particular there is $1 \in \mathbb{K}$ for which $1\boldsymbol{v} = \boldsymbol{v}$
(iv) $a(\boldsymbol{v} + \boldsymbol{w}) = a\boldsymbol{v} + a\boldsymbol{w}$
(v) $(ab)\boldsymbol{v} = a(b\boldsymbol{v})$
(vi) There is a unique zero element $\boldsymbol{0} \in V$ such that $\boldsymbol{v} + \boldsymbol{0} = \boldsymbol{v}$
We refer to the elements of a vector space as *vectors* [5].

For the remainder of the text, we shall take $\mathbb{K}$ to be the real numbers $\mathbb{R}$. Now let V be a vector space of dimension n spanned by a set of *basis vectors* $\{\boldsymbol{e}_i\}$, $i \in \{1, ..., n\}$. Denoting the i:th component of $\boldsymbol{v} \in V$ as v^i and adopting the *Einstein summation convention* of implicitly summing over all indices that appear both up and down, we may write $\boldsymbol{v}$ in *component notation* as

$$\boldsymbol{v} = \sum_{i=1}^{n} v^i \boldsymbol{e}_i \equiv v^i \boldsymbol{e}_i \quad \text{for } v^i \in \mathbb{R}, \quad \boldsymbol{e}_i \in V$$

In ordinary calculus the tangent of a function $f(x) : \mathbb{R} \to \mathbb{R}$ at $x = x_0$ is obtained by the first derivative $(df/dx)_{x_0}$. In a similar manner, the directional derivatives of a function f evaluated at a point x on a manifold M give the tangent vectors of M at x. By assigning a vector to each point $x \in M$, we construct a *vector field*, whose action on a real-valued function f is given by

$$\boldsymbol{v}_x(f) = v^i \left(\frac{\partial f}{\partial x^i} \right)_x \equiv v^i (\partial_i f)_x, \tag{3.1}$$

where the subscript x denotes that the partial derivative is evaluated pointwise. The set of all vectors tangent to M at x constitutes a vector space known as the *tangent space* $T_x M$. It is customary to omit the subscript in Eq. (3.1) and those alike for clarity, as calculations tend to contain already appreciable amounts of sub- and superscripts. Nevertheless, the reader ought not to forget that

T_xM only exists at a certain point $x \in M$. This has the consequence that we may not simply translate a vector $\boldsymbol{v} \in T_xM$ around the manifold, but have to transform it into a whole another tangent space that is defined at the destination point. To do calculations involving vectors from different tangent spaces at separate points of a manifold will require some additional machinery, which we will eventually get to. As Eq. (3.1) suggests, we can identify the basis vectors of T_xM with the partial derivatives

$$\{\boldsymbol{e}_i\} \to \partial_i,\ \boldsymbol{v} = v^i\boldsymbol{e}_i \to \boldsymbol{v} = v^i\partial_i. \tag{3.2}$$

To produce numbers out of vectors we use *dual vectors*, specified by the following definitions.

Definition 3.2. Let V be a vector space. The **dual vector space** (dual space) V^* of dimension $\dim(V^*) = \dim(V)$ is the vector space of linear functions $\boldsymbol{v}^* : V \to \mathbb{R}$ that map a vector $\boldsymbol{v} \in V$ to a real number $\boldsymbol{v}^*(\boldsymbol{v}) \in \mathbb{R}$. As a linear function, a dual vector $\boldsymbol{v}^* \in V^*$ has the property

$$\boldsymbol{v}^*(a\boldsymbol{v} + b\boldsymbol{w}) = a\boldsymbol{v}^*(\boldsymbol{v}) + b\boldsymbol{v}^*(\boldsymbol{w}) \text{ for all } \boldsymbol{v}, \boldsymbol{w} \in V \text{ and } a, b \in \mathbb{R}.$$

The reader should notice that dual vectors are usually denoted by greek letters in mathematics literature. We will however reserve greek letters for other purposes and denote dual vectors by an asterisk. This also emphasizes the one–to–one correspondence between vectors and their duals in later sections.

Definition 3.3. Let V be a vector space spanned by $\{\boldsymbol{e}_i\}$ and V^* its dual space spanned by $\{\boldsymbol{e}^j\}$; $i, j \in \{1, ..., \dim(V)\}$. The action of the basis dual vectors of V^* on the basis vectors of V is defined to be equal to the Kroenecker delta symbol:

$$\boldsymbol{e}^i(\boldsymbol{e}_j) \equiv \delta^i_j \equiv \begin{cases} 1, i \neq j \\ 1, i = j \end{cases} \quad .$$

Let us continue working with the spaces and bases of definition 3.3. Sticking to the Einstein summation convention, we indicate the components of a dual vector $\boldsymbol{w}^* \in V^*$ by a lower index so that $\boldsymbol{w}^* = w_i\boldsymbol{e}^i$. The operation $\boldsymbol{w}^*(\boldsymbol{v}), \boldsymbol{v} \in V$ becomes

$$\boldsymbol{w}^*(\boldsymbol{v}) = \boldsymbol{w}^*(v^i\boldsymbol{e}_i) = v^i\boldsymbol{w}^*(\boldsymbol{e}_i) = v^iw_j\boldsymbol{e}^j(\boldsymbol{e}_i) = v^iw_i$$

The dual space of T_xM is the *cotangent space* T^*_xM and the linear functions mapping the tangent vectors to $\mathbb{R}$ are known as *covectors*. Due to any two vector spaces of the same dimension being isomorphic, we have that the tangent and cotangent spaces are dual to each other [5]:

$$(T^*_xM)^* = T_xM.$$

To make an identification similar to Eq. (3.2) for the dual basis, we turn to *differentials*.

Definition 3.4. Let M be a C^∞ manifold. The action of a **differential** $df \in T^*_xM$ at point $x \in M$ on a vector field $\boldsymbol{v} \in T_xM$ is defined to be [1]

$$df(\boldsymbol{v}) \equiv \boldsymbol{v}(f).$$

Using the basis $\{\partial_i\}$ with local coordinates $\{x^i\}$, we continue to write $df(\boldsymbol{v}) = v^i \partial_i f$. This indicates a dual correspondence between differentials and partial derivatives, hence suggesting that we identify the basis dual vectors with the differentials of the coordinates [1, 2]

$$\{\boldsymbol{e}^i\} \rightarrow \{dx^i\}.$$

To further illustrate the idea, consider the total differential of a function f in local coordinates $\{x^i\}$. It is usually written $df = (\partial_i f)dx^i$ and, by definition 3.4, a differential acts on a vector to produce a real number. A general $\boldsymbol{v}^*$ should then behave the same way, but with some functions v_i to multiply the dx^i. Notice that using the coordinate basis $\{\partial_i\}$ and the dual basis $\{dx^i\}$ also gives the expected result for the action of a dual vector on a vector:

$$\boldsymbol{v}^*(\boldsymbol{v}) = v_i dx^i (v^j \partial_j) = v_i v^j dx^i(\partial_j) = v_i v^j \frac{\partial x^i}{\partial x^j} = v_i v^j \delta^i_j = v_i v^i.$$

The notion of linear maps is generalized to *multilinear maps* $\boldsymbol{F} : V \times ... \times V \rightarrow \mathbb{R}$ by allowing $\boldsymbol{F}$ to take $k \in \mathbb{N}$ input arguments and demanding it to be linear in each one of them. We will also need multilinear maps $\boldsymbol{G} : V^* \times ... \times V^* \rightarrow \mathbb{R}$ that take $l \in \mathbb{N}$ dual vectors to $\mathbb{R}$. To get the most general multilinear functions we need to combine the features of both, giving rise to *tensors*. Along the lines of [2], we state the following definitions.

Definition 3.5. Let V be a vector space and V^* its dual. A (k,l)**-tensor** is a multilinear map $\boldsymbol{T} : V^* \times ... \times V^* \times V \times ... \times V \rightarrow \mathbb{R}$, taking $k \in \mathbb{N}$ dual vectors and $l \in \mathbb{N}$ vectors as input arguments. The *rank* of the tensor is the total amount of arguments $k + l$. We will denote tensors by bold symbols, unless some other special symbol is especially given.

Instead of (k,l), some texts refer to k-covariant l-contravariant tensors by $\binom{k}{l}$, portraying the amounts of upper and lower indices. We have chosen the convention of [6].

Definition 3.6. The **tensor product** of a (k,l)-tensor $\boldsymbol{T}_1$ and an (m,n)-tensor $\boldsymbol{T}_2$ results in a $(k+m, l+n)$-tensor $\boldsymbol{T}_1 \otimes \boldsymbol{T}_2$, whose action on a set of dual vectors $\{\boldsymbol{w}^*_1, ..., \boldsymbol{w}^*_{k+m}\}$ and a set of vectors $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_{l+n}\}$ is given by

$$(T_1 \otimes T_2)(\boldsymbol{w}^*_1, ..., \boldsymbol{w}^*_{k+m}, \boldsymbol{v}_1, ..., \boldsymbol{v}_{l+n}) = T_1(\boldsymbol{w}^*_1, ..., \boldsymbol{w}^*_k, \boldsymbol{v}_1, ..., \boldsymbol{v}_l) T_2(\boldsymbol{w}^*_{k+1}, ..., \boldsymbol{w}^*_{k+m}, \boldsymbol{v}_{l+1}, ..., \boldsymbol{v}_{l+n}).$$

Definition 3.7. Suppose the vector space V has a basis $\{\boldsymbol{e}_i\}$ and its dual V^* has a dual basis $\{\boldsymbol{e}^i\}$, $i \in 1, ..., n$. The **tensor space** $T^k_l(V)$ of (k,l)-tensors is then spanned by

$$\{\boldsymbol{e}_{i_1} \otimes ... \otimes \boldsymbol{e}_{i_k} \otimes \boldsymbol{e}^{j_1} \otimes ... \otimes \boldsymbol{e}^{j_l}\}, \quad i_1, ..., i_k, j_1, ..., j_l \in \{1, ..., n\}.$$

The action of the basis tensor on the basis vectors and basis dual vectors is specified by

$$(\boldsymbol{e}_{i_1} \otimes ... \otimes \boldsymbol{e}_{i_k} \otimes \boldsymbol{e}^{j_1} \otimes ... \otimes \boldsymbol{e}^{j_l})(\boldsymbol{e}^{a_1}, ..., \boldsymbol{e}^{a_k}, \boldsymbol{e}_{b_1}, ..., \boldsymbol{e}_{b_l}) \equiv \delta^{a_1}_{i_1} ... \delta^{a_k}_{i_k} \delta^{j_1}_{b_1} ... \delta^{j_l}_{b_l}. \tag{3.3}$$

We generalize the action of a basis tensor on tensors formed via the tensor product of some basis vectors and / or covectors by replacing any comma in the argument section of Eq. (3.3) by $\otimes$.

Notice that in definition 3.7 the index pairs on the deltas arise as if the dual basis vector corresponding to the upper index was always acting on a basis vector corresponding to the lower index, regardless of which one appears as an argument in the first place. With these definitions, we can generalize vectors to $(1,0)$-tensors and dual vectors to $(0,1)$-tensors[2]. Using definition 3.7, we may express tensors in terms of their components by writing

$$\boldsymbol{T} = T^{i_1...i_k}{}_{j_1...j_l} \boldsymbol{e}_{i_1} \otimes ... \otimes \boldsymbol{e}_{i_k} \otimes \boldsymbol{e}^{j_1} \otimes ... \otimes \boldsymbol{e}^{j_l},$$

where we drop the boldface for tensor components, as with vectors and dual vectors. We may also write $\boldsymbol{T}$ using the coordinate basis $\{\partial_i\}$ and dual basis $\{dx^i\}$ as

$$\boldsymbol{T} = T^{i_1...i_k}{}_{j_1...j_l} \partial_{i_1} \otimes ... \otimes \partial_{i_k} \otimes dx^{j_1} \otimes ... \otimes dx^{j_l}.$$

The action of a (k,l)-tensor on $\{\boldsymbol{w}_1^*, ..., \boldsymbol{w}_k^*\} \in V^*$ and $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_l\} \in V$ then becomes

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{w}_1^*, ..., \boldsymbol{w}_k^*, \boldsymbol{v}_1, ..., \boldsymbol{v}_l) &= T^{i_1...i_k}{}_{j_1...j_l} (\partial_{i_1} \otimes ... \otimes \partial_{i_k} \otimes dx^{j_1} \otimes ... \otimes dx^{j_l})(\boldsymbol{w}_1^*, ..., \boldsymbol{w}_k^*, \boldsymbol{v}_1, ..., \boldsymbol{v}_l) \\ &= T^{i_1...i_k}{}_{j_1...j_l} w_{1i_1} ... w_{ki_k} v_1^{j_1} ... v_l^{j_l}, \end{aligned}$$

Note that here for instance w_{1i_1} is the i_1:th component of the dual vector $\boldsymbol{w}_1$, not a component of some rank 2 tensor. Whenever such notation is used, one should keep in mind whether or not some subscript is actually to be interpreted as a part of the symbol used for denoting the dual vector and not an index that could be summed over. Further, the lower indices in the coefficient $T^{i_1...i_k}{}_{j_1...j_l}$ are called *covariant* and the upper indices *contravariant*. The convention is in analogy with the behaviour of certain mathematical objects under a change in the units of length: a covariant index is summed together with a differential dx^i and a $(0,1)$-tensor is thus also called a covariant vector. Notice that changing the units of length varies a covariant vector directly. However, a contravariant index is summed with $\partial_i = \partial/\partial x^i$. It follows that a $(1,0)$-tensor or a *contra*variant vector then varies *inversely* with respect to a change in the units of length.

The order of the covariant and contravariant indices is the same as that of the tensor's input arguments. The arguments' nature as either vectors or dual vectors is revealed by whether the corresponding index is a covariant or a contravariant one. The amounts of the two kinds of indices give the covariant and contravariant *orders* of the tensor, respectively. The components thus carry essentially the same information as writing the tensor explicitly as a multilinear map and specifying the arguments it is to act on. Therefore it is customary in physics literature to write tensor equations in terms of their components.

With a choice of basis ∂_i and dual basis dx^i, $i \in \{1, ..., n\}$ we may represent a vector $\boldsymbol{v} \in T_xM$ by listing its components as in

$$\boldsymbol{v} = (v^1, ..., v^n) \doteq v^i,$$

where we use the dotted equality as a reminder that v^i stands for the ith component of $\boldsymbol{v}$; it is *not* the same mathematical object as the whole vector. Instead, v^i is one of the numbers $v^1, ..., v^n$ at a time, depending on the choice of i. We may also represent covectors in the same way. As the tangent and cotangent spaces are of the same dimension n, we may represent $(2,0)$, $(1,1)$ and $(0,2)$

[2]It is because of this that we have chosen to denote them by bold symbols just as any other tensors. A $(0,0)$-tensor would be a scalar.

tensors as $n \times n$ matrices. For instance, a $(0,2)$-tensor $\boldsymbol{T}$ is customarily written as

$$\boldsymbol{T} = \begin{bmatrix} T_{11} & \dots & T_{1n} \\ \vdots & \ddots & \vdots \\ T_{n1} & \dots & T_{nn} \end{bmatrix} \doteq T_{ij},$$

so that the first index corresponds to a row and the second to a column, making T_{ij} a single component of the matrix at a time. Observe that there is a risk of ambiguity in the first equality. It is good practice to somehow specify what kind of arguments $\boldsymbol{T}$ takes when writing such equations. $\boldsymbol{T}$ could be a $(2,0)$, $(0,2)$ or $(1,1)$-tensor, all of which are different objects in the sense that their matrix representations will not agree in general. This is encoded in component notation by recognizing that $T_{ij} \neq T^{ij}$, $T_{ij} \neq T^{i}{}_{j}$, $T_{ij} \neq T_{i}{}^{j}$, $T^{ij} \neq T_{i}{}^{j}$, $T^{ij} \neq T^{i}{}_{j}$, $T_{i}{}^{j} \neq T^{i}{}_{j}$.

An equation relating tensor components, for instance

$$B_{i}{}^{k} C_{kj} + D_{jik} v^{k} + v_j w_i = A_{ij}$$

is to be interpreted so that each term ought to have the same remaining *free indices* (i and j above) after all Einstein summations over *dummy indices* (k above; however, the choice of symbol for a dummy index is arbitrary) have been carried out. That is, the co- and contravariant orders of each term must match for the equation to make sense if we are ever to turn the equation back into a tensor relation instead of working with the components that are just numbers. The above example also serves to demonstrate some of the possible ways of ending up with two free lower indices. Summing over an index shared by two tensors is called a *contraction*, though a tensor with enough co- and contravariant indices may also be contracted with itself. In particular, the *trace* of a $(1,1)$ tensor is $T^{i}{}_{i} = T_{i}{}^{i} = T$, where the last T is the resulting scalar value and not to be mistaken for the whole tensor.

At this point, there is an important note on coordinate independecy to be made. Even if entering the component notation takes the calculation to a local coordinate system, a notation such as $\boldsymbol{T}(\boldsymbol{w}_1^*, ..., \boldsymbol{w}_k^*, \boldsymbol{v}_1, ..., \boldsymbol{v}_l)$ does not require a choice of basis. This is because tensors itself are *coordinate free*. If a tensor relation holds in one coordinate system, it will hold in any other one as well. The real numbers produced by tensor calculations are invariant *scalar values*. To illustrate this, suppose both $\{\boldsymbol{e}_i\}$ and $\{\boldsymbol{e}'_i\}$ with $i \in \{1, .., \dim(V)\}$ span a vector space V. Let $\{x^i\}$ and $\{x'^i\}$ be the corresponding local coordinates. Then a vector $\boldsymbol{v} \in V$ may be written $\boldsymbol{v} = v^i \boldsymbol{e}_i = v'^i \boldsymbol{e}'_i$ so that the vector as a mathematical object is unchanged in a coordinate transformation $\{\boldsymbol{e}_i\} \to \{\boldsymbol{e}'_i\}$, despite the components v^i change into v'^i. In fact, the components change precisely as they should to make the coordinate invariance of the whole vector itself possible: by the elements of the Jacobian matrix

$$v'^i = \frac{\partial x'^i}{\partial x^j} v^j. \tag{3.4}$$

Similar reasoning holds also for more general tensors. Strictly speaking, the procedure for classifying whether or not a map is induced by a tensor field is to demand the properties of a tensor characterization lemma given in e.g. [2]. For our purposes it is however adequate to be aware of the following lemma, extending the idea of equation (3.4) to tensors of arbitrary covariant and contravariant orders.

Lemma 3.8. ***(Tensor transformation law)*** *The components of a tensor change under a coordinate transformation as*

$$T'^{i_1...i_k}{}_{j_1...j_l} = \frac{\partial x'^{i_1}}{\partial x^{a_1}} ... \frac{\partial x'^{i_k}}{\partial x^{a_k}} \frac{\partial x^{b_1}}{\partial x'^{j_1}} ... \frac{\partial x^{b_l}}{\partial x'^{j_l}} T^{a_1...a_k}{}_{b_1...b_l}.$$

Proof. By the chain rule, we have that

$$dx'^i = \frac{\partial x'^i}{\partial x^j} dx^j \quad \text{and} \quad \partial'_i = \frac{\partial}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j} = \frac{\partial x^j}{\partial x'^i} \partial_j$$

Lemma 3.8 is then obtained by inserting these to $\boldsymbol{T} = T'^{i_1...i_k}{}_{j_1...j_l} \partial'_{i_1} \otimes ... \otimes \partial'_{i_k} \otimes dx'^{j_1} \otimes ... \otimes dx'^{j_l}$ and setting the result equal to $\boldsymbol{T} = T^{i_1...i_k}{}_{j_1...j_l} \partial_{i_1} \otimes ... \otimes \partial_{i_k} \otimes dx^{j_1} \otimes ... \otimes dx^{j_l}$, when

$$\left(\frac{\partial x'^i}{\partial x^j} \right)^{-1} = \frac{\partial x^j}{\partial x'^i} \quad \text{and} \quad \left(\frac{\partial x^j}{\partial x'^i} \right)^{-1} = \frac{\partial x'^i}{\partial x^j}.$$

□

4. From a manifold to another

The tangent, cotangent and tensor spaces on a manifold are in fact manifolds by themselves. They are however defined pointwise, so we need to define some additional structure to be able to consider *fields* of vectors, covectors and tensors extending over a manifold. This requires us to define certain maps, starting with *diffeomorphisms* and smooth maps to move between manifolds. We proceed here mostly as in [1].

Definition 4.1. **(Smooth maps of manifolds)** Let M and N be C^∞ manifolds of dimensions m and n, respectively. Suppose we have the charts (U, φ) on M and (V, ψ) on N such that the neighbourhood U is about a point $x \in M$ and V about $y \in N$. A map $F : M \to N$ is C^∞ at $x \in M$ if the composition $\psi \circ F \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is C^∞ at $\varphi(x) \in \mathbb{R}^m$. If this holds for all $x \in M$, we call F C^∞ on M.

Definition 4.2. Let M and N be smooth manifolds. A **diffeomorphism** between manifolds is the C^∞ bijection $F : N \to M$ with a C^∞ inverse $F^{-1} : M \to N$. We call M and N **diffeomorphic** if their tangent spaces at the corresponding points mapped to each other by the bijection are isomorphic [1].

To combine all tangent spaces at different points of the manifold we need the concept of *vector bundles.*

Definition 4.3. Let E and M be smooth manifolds. A smooth **vector bundle** denoted by $\pi : E \to M$ is the triple (E, M, π), where π is a surjective map known as the *projection.* Now let

$x \in M$ be a point in a neighbourhood U. We require π to have the following properties [1, 2].

(i) The *fiber* of E over x is the set $E_x \equiv \pi^{-1}(x)$ with vector space structure.

(ii) For every $x \in M$ there is (U, φ) so that the *local trivialization* $\varphi : \pi^{-1}(U) \to U \times \mathbb{R}^n$ is a diffeomorphism and the below diagram commutes.

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "\varphi"] \arrow[dr, "\pi"'] & U \times R^n \arrow[d, "\pi'"] \\ & U \end{tikzcd}$$

(iii) By restricting φ to a fiber E_x we get a vector space isomorphism $\varphi_x : E_x \to \{x\} \times \mathbb{R}^n$.

The vector bundle that will be of the most use to us is the *tangent bundle* $(TM, M, \pi : TM \to M)$. For brevity, it is often denoted simply by just TM, which is the disjoint union of all tangent spaces T_xM at all points $x \in M$ [2]:

$$TM \equiv \coprod_{x \in M} T_xM.$$

In definition 4.3, E is usually called the *total space* and M the *base*. Loosely speaking, the meaning of this may be illustrated if we consider the tangent bundle of a manifold M that can be embedded into some higher dimensional space. M then fills some part of the space, whereas all the tangent vectors in TM occupy the surroundings of M. Another useful vector bundle is the *normal bundle*. There we have the disjoint union NM of all the vector spaces associated with some point on M, on which lie the vectors whose projection from the total space to the base manifold M returns zero.

Now let $\pi : E \to M$ be a vector bundle over M. A *section* S of E is a map $S : M \to E$ such that $\pi \circ S$ is the identity map on M. The section is smooth if S is smooth as a map between manifolds. The importance of tangent bundles and smooth sections is that they give us a way to assign a tangent vector to each point of a manifold: a smooth *vector field* on M is the smooth section $S_T(M)$ of the tangent bundle TM. Similarly to the tangent bundle, the *cotangent bundle* T^*M is the disjoint union of cotangent spaces $T^*_xM = (T_xM)^*$. *Covector fields* or *one-forms* (1-forms) are then formed analogously as the smooth sections of the cotangent bundle. In general, the *tensor bundle* T^k_lM of (k, l)-tensors is the disjoint union of all tensor spaces at different points x of a manifold M

$$T^k_lM \equiv \coprod_{x \in M} T^k_l(T_xM).$$

As a logical continuum to the previous cases, a *tensor field* is then a smooth section of some tensor bundle T^k_lM [1]. Lastly, we consider what happens to vectors and 1-forms under a C^∞ map of two C^∞ manifolds. Let us begin with vectors.

Definition 4.4. Let M and N be C^∞ manifolds. The *differential* of the C^∞ map of manifolds $\varphi : M \to N$ at $x \in M$ is the **pushforward** $\varphi_* : T_xM \to T_{\varphi(x)}N$ of a vector in T_xM [1].

Notice that the pushforward is defined pointwise ergo for the vectors at a single point. Now let us extend the notion to vector fields, supposing in addition to the objects of definition 4.4 that $\boldsymbol{v}$ is a vector field on M. We do this by demanding that φ is a diffeomorphism. The bijective nature of φ then ensures that $(\varphi_*(\boldsymbol{v}))_{\varphi(x)} = \varphi_{*x}(\boldsymbol{v}_x)$ and that the pushforward is defined everywhere on M [1].

The dual of the differential φ_* is the *codifferential*. The map φ can be used to push a vector in T_xM to $T_{\varphi(x)}N$ using the differential, whereas the codifferential serves to pull 1-forms back from N to M.

Definition 4.5. Let $\varphi : M \to N$ be a C^∞ map of the C^∞ manifolds M and N with the differential $\varphi_* : T_xM \to T_{\varphi(x)N}$. The codifferential $\varphi^* : T^*_{\varphi(x)}N \to T^*_xM$ maps a 1-form on N to a 1-form on M. This is called the **pullback** of a 1-form and denoted $\varphi^*(\boldsymbol{w}^*_{\varphi(x)}) = (\varphi^*\boldsymbol{w}^*)_x$. For a vector $\boldsymbol{v}_x \in T_xM$ at $x \in M$ pushed forward by φ_* and a 1-form $\boldsymbol{w}^*$ on N, we have $(\varphi^*\boldsymbol{w}^*)_x(\boldsymbol{v}_x) = \boldsymbol{w}^*_{\varphi(x)}(\varphi_*(\boldsymbol{v}_x))$ [1].

5. The Riemannian metric

To be able to do geometry, we need a way to measure angles and distances on the manifold. This can be done by using the *metric tensor field*. In a sense, it tells us how space is behaving in different directions as the field exists everywhere on the manifold and is used to define certain essential operations in Riemannian geometry.

Definition 5.1. Let M be a smooth manifold and $\boldsymbol{v}, \boldsymbol{w} \in T_xM$. The **Riemannian metric** is a symmetric rank 2 tensor field $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{w}, \boldsymbol{v})$. It is positive definite, so that $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) > 0$ for any $\boldsymbol{v} > \boldsymbol{0}$ and it determines the inner product (dot product) of two vectors in the tangent space T_xM as $\langle \boldsymbol{v}, \boldsymbol{w} \rangle \equiv \boldsymbol{v} \cdot \boldsymbol{w} \equiv \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w})$. In local coordinates $\{x^i\}$ we may also write $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) = g_{ij}v^iw^j$. A smooth manifold with a Riemannian metric is a **Riemannian manifold** [2].

A frequently occuring example of a Riemannian manifold is the Euclidean space $\mathbb{R}^n$, for which the usual matrix representation of the metric tensor is the identity matrix $\text{diag}(1, ..., 1)$. Hence the components $g_{ij} = \delta_{ij}$. Since any Riemannian manifold is locally Euclidean, we can always choose a local coordinate frame where the representation of g_{ij} is diagonal, or even agrees with the identity matrix [2].

Let M be a Riemannian manifold with the Riemannian metric $\boldsymbol{g}$. Using local coordinates $\{x^i\}$, we can calculate ds^2, the squared length of an infinitesimal line-element[3]

$$ds^2 = g_{ij}dx^idx^j,$$

where the power of two in ds^2 is not to be mistaken for an index. Customarily, whenever a risk of confusion should arise, powers can be indicated by braces as in $(x^i)^2$. In some local frame, the components of the metric tensor are the inner products of the basis vectors [2]

$$g_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle. \tag{5.1}$$

We can proceed further to calculate the length s of a curve given by the coordinates $\{x^i(\tau)\}$ by the integration

$$s = \int \sqrt{ds^2} = \int \sqrt{g_{ij}dx^idx^j} = \int \sqrt{g_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}}d\tau$$

Although the above notation is naive, it serves to build intuition to the coordinate free expression obtained if we suppose the curve has a tangent vector $\boldsymbol{t}$ with the components $\dot{x}^i$

$$s = \int \sqrt{\boldsymbol{g}\left(\boldsymbol{t}, \boldsymbol{t}\right)}d\tau. \tag{5.2}$$

[3]Notice that for $\mathbb{R}^n$ this reduces to the Pythagorean theorem: $ds^2 = (dx^1)^2 + ... + (dx^n)^2$

The length or norm $|\boldsymbol{v}|$ of a vector $\boldsymbol{v} \in T_xM$ is also given by the metric via the inner product $|\boldsymbol{v}| = \langle \boldsymbol{v}, \boldsymbol{v} \rangle^{1/2}$. The angle θ between two vectors $\boldsymbol{v}, \boldsymbol{w} \in T_xM$ with non-zero lengths $|\boldsymbol{v}| \neq 0$, $|\boldsymbol{w}| \neq 0$ is then obtained from

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = |\boldsymbol{v}||\boldsymbol{w}| \cos(\theta)$$
$$\theta = \arccos\left(\frac{\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w})}{|\boldsymbol{v}||\boldsymbol{w}|}\right).$$

The Riemannian metric g also serves to map a vector $\boldsymbol{v}$ in the tangent space T_xM of a Riemannian manifold M at $x \in M$ into its dual on the cotangent space T_x^*M by the *flat* or *lowering* operation

$$\boldsymbol{v}^\flat = (g_{ij}v^j)dx^i = v_i dx^i.$$

There is a corresponding operation to turn a 1-form $\boldsymbol{v}^* \in M$ into its dual, the vector $\boldsymbol{v} \in T_xM$. For this we however first need to define the inverse of $\boldsymbol{g}$.

Definition 5.2. Let $\boldsymbol{g}$ be a Riemannian metric on a Riemannian manifold M. The components of the **inverse metric** $\boldsymbol{g}^{-1}$ are specified by $g_{ij}g^{jk} \equiv \delta_i^k$.

Now the *sharp* or *raising* operation can be stated as [2].

$$\boldsymbol{v}^{*\sharp} = (g^{ij}v_j)\partial_i.$$

Following this logic, we may check that the subsequent application of components of $(\boldsymbol{v}^*)^* \in (T_x^*M)^* = T_xM$ are again the components of $\boldsymbol{v} \in T_xM$, as given by $g^{jk}g_{ij}v^i = \delta_i^k v^i = v^k$. A similar result also holds for covectors. The raising and lowering operations also enable us to contract a (k,l)-tensor with itself. We may then *trace out* any two indices of a (k,l)-tensor to produce a $(k-2,l)$-tensor with the components

$$g_{i_n i_m} T^{i_1 \dots i_n \dots i_m \dots i_k}{}_{j_1 \dots j_l} = T^{i_1 \dots i_n \dots i_{m-1}}{}_{i_n}{}^{i_{m+1} \dots i_k}{}_{j_1 \dots j_l}. = T^{i_1 \dots i_{n-1} i_{n+1} \dots i_{m-1} i_{m+1} \dots i_k}{}_{j_1 \dots j_l}.$$

An analogous procedure utilizing the inverse metric tensor components would yield the components of a $(k, l-2)$-tensor.

6. Curvature and geodesics

A curve is a one-dimensional geometrical object homeomorphic to the set of real numbers $\mathbb{R}$. It can be embedded into a higher dimensional space $\mathbb{R}^n$ when $n > 1$, but this doesn't change the dimensionality of the curve in an *intrinsic* sense: any being restricted to live on the curve would still experience a world with only one dimension. Analogously, we consider surfaces to be intrinsically two-dimensional objects. In general, an n-manifold is intrinsically n-dimensional. In a similar fashion, we should not consider a manifold with $n \geq 2$ dimensions to be curved just because it might *appear* to be curved by how it has been set up in some higher-dimensional space. Such *extrinsic curvature* does not affect how a being living on the manifold measures angles and distances, which are precisely the properties we should look at when examining *intrinsic curvature*. For instance, albeit its extrinsically curved appearance, the surface of a cylinder is not intrinsically curved: measurements of angles and distances are performed as on any other two-dimensional Euclidean space.

In flat Euclidean space one may move any vector to "start" at an arbitrary point. This is because the tangent spaces happen to be the same at each point. This is however no more a property of general curved manifolds. Thus any operation involving the comparison of several vectors belonging to different pointwise-defined tangent spaces will require a map to *connect* the different tangent spaces.

Definition 6.1. Let E and M be smooth manifolds, $\pi : E \to M$ a vector bundle, $S_E(M) \equiv S_E$ the smooth sections of E and $S_T(U)$ the smooth sections of the tangent bundle in some neighbourhood U of M. A **connection** is the map $\nabla : S_T(U) \times S_E \to S_E$ with the following defining properties [2] for all $a, b \in \mathbb{R}; f, h \in M$.

(i) $\nabla_{f\boldsymbol{v}+h\boldsymbol{w}}\boldsymbol{u} = f\nabla_{\boldsymbol{v}}\boldsymbol{u} + h\nabla_{\boldsymbol{w}}\boldsymbol{u}$

(ii) $\nabla_{\boldsymbol{v}}(a\boldsymbol{w} + b\boldsymbol{u}) = a\nabla_{\boldsymbol{v}}\boldsymbol{w} + b\nabla_{\boldsymbol{v}}\boldsymbol{u}$

(iii) $\nabla_{\boldsymbol{v}}(f\boldsymbol{w}) = (\boldsymbol{v}(f))\boldsymbol{w} + f\nabla_{\boldsymbol{v}}\boldsymbol{w}$

From now on we shall work with *linear* connections by considering vector bundles that are the manifold's tangent bundles. In a local frame, the covariant derivative of a basis vector $\boldsymbol{e}_i$ is then given by the expansion

$$\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j \equiv \Gamma^k_{ij}\boldsymbol{e}_k,$$

where the coefficients Γ^i_{jk} are known as *Christoffel symbols* or *connection coefficients*. Despite the appearance, they are not tensor components, as they do not transform by the requirement of lemma 3.8. The operation $\nabla_{\boldsymbol{v}}\boldsymbol{w}$ associated with a connection ∇ is the **covariant derivative** of $\boldsymbol{w}$ in a direction specified by $\boldsymbol{v}$. It can be written as

$$\begin{aligned}
\nabla_{\boldsymbol{v}}\boldsymbol{w} &= \nabla_{v^i\boldsymbol{e}_i}(w^j\boldsymbol{e}_j) \\
&= v^i\nabla_{\boldsymbol{e}_i}(w^j\boldsymbol{e}_j) \\
&= v^i\left(\boldsymbol{e}_i(w^j)\boldsymbol{e}_j + w^j\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j\right) \\
&= v^i\left(\boldsymbol{e}_i(w^j)\boldsymbol{e}_j + w^j\Gamma^k_{ij}\boldsymbol{e}_k\right) \\
&= v^i\left(\boldsymbol{e}_i(w^j) + \Gamma^j_{ik}w^k\right)\boldsymbol{e}_j \\
&\equiv v^i(\nabla_i w^j)\boldsymbol{e}_j
\end{aligned}$$

so that we have denoted the *covariant derivative acting on the components* w^i as $\nabla_i w^j$. These are also the components of a tensor $\nabla\boldsymbol{w}$ whose action on a vector $\boldsymbol{v}$ is given by $\nabla\boldsymbol{w}(\boldsymbol{v}) \equiv \nabla_{\boldsymbol{v}}\boldsymbol{w}$. Let us now choose to work in a coordinate basis where we make the identification $\{\boldsymbol{e}_i\} \to \{\partial_i\}$. We then have $\nabla\boldsymbol{w} = (\nabla_i w^j)dx^i \otimes \partial_j$ and the components become

$$\nabla_i w^j \equiv \partial_i w^j + \Gamma^j_{ik}w^k.$$

The action of ∇_i on the components w_j differs by the sign of the Christoffel symbol [6]

$$\nabla_i w_j \equiv \partial_i w_j - \Gamma^k_{ij}w_k.$$

In general, the covariant derivative acting on the components of a (k, l)-tensor $\boldsymbol{T}$ is given by

$$\nabla_a T^{i_1\ldots i_k}{}_{j_1\ldots j_l} \equiv \partial_a T^{i_1\ldots i_k}{}_{j_1\ldots j_l} + \sum_{n=1}^{k}\Gamma^{i_n}_{ab}T^{i_1\ldots b\ldots i_k}{}_{j_1\ldots j_l} - \sum_{m=1}^{l}\Gamma^b_{aj_m}T^{i_1\ldots i_k}{}_{j_1\ldots b\ldots j_l},$$

where the last two terms are to be understood such that in each term of the sums b replaces the index i_n or j_n, which has been moved to the Christoffel symbol [6].

For our examinations in the following chapters, a particularly interesting case of covariant derivatives will be those along a curve $\gamma(\tau)$ on a Riemannian manifold M. Suppose $\boldsymbol{v}$ is an *extensible vector field*, i.e. one that is not restricted to live on the curve but for which we can have an extension $\tilde{\boldsymbol{v}}$ covering also other parts of the manifold. The covariant derivative of a vector $\boldsymbol{v}$ along $\gamma(\tau)$ is denoted by

$$D_\tau \boldsymbol{v} = \nabla_{\boldsymbol{t}} \tilde{\boldsymbol{v}},$$

where $\boldsymbol{t} = \left(\frac{d}{d\tau}\gamma^i(\tau)\right)\partial_i \equiv \dot{\gamma}^i(\tau)\partial_i$ is the tangent vector of $\gamma(\tau)$ and often labeled the velocity on the curve. Here we will use *affine parametrizations*, meaning that the curve lenght s is linear in the curve parameter τ such that $s = a\tau + b$ for some constants $a, b \in \mathbb{R}$. D_τ also has the following properties [2]: it is linear in its arguments

$$D_\tau(a\boldsymbol{v} + b\boldsymbol{w}) = aD_\tau \boldsymbol{v} + bD_\tau \boldsymbol{w}$$

and satisfies the product rule

$$D_\tau(f\boldsymbol{v}) = \left(\frac{df}{d\tau}\right)\boldsymbol{v} + fD_\tau \boldsymbol{v}.$$

When transporting a vector $\boldsymbol{v}$ along a curve $\gamma(\tau)$, we say that it is *parallel transported* if the *autoparallelity requirement*

$$D_\tau \boldsymbol{v} = \boldsymbol{0} \tag{6.1}$$

holds. Now let us consider a point x on a curve γ on a location where the curve parameter has some value τ. The derivatives of the point's coordinates $\{x^i(\tau)\}$ with respect to τ then let us write the covariant derivative of the tangent vector $\boldsymbol{t}$ along its own direction by using the components $\dot{x}^i\partial_i$. For the curves that parallel transport their own tangent vector, the autoparallelity requirement yields

$$\begin{aligned} D_\tau \boldsymbol{t} = \boldsymbol{0} \;\Leftrightarrow\; \left(\frac{d\dot{x}^j}{d\tau}\right)\partial_j + \dot{x}^j\dot{x}^i\nabla_{\partial_i}\partial_j &= 0 \\ (\ddot{x}^j + \dot{x}^k\dot{x}^i\Gamma^j_{ik})\partial_k &= 0 \end{aligned}$$

Such curves are called *geodesics* and must be parametrized by an affine parameter. The components of the result in the above calculation give the *geodesic equation*

$$\ddot{x}^j + \Gamma^j_{ik}\dot{x}^i\dot{x}^k = 0.$$

We now turn to two properties that let us determine a unique connection on a Riemannian or pseudo-Riemannian manifold. The first one is *metric compatibility.*

Definition 6.2. Let M be a Riemannian manifold with a metric tensor $\boldsymbol{g}$ and arbitrary vector fields $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in TM$. The connection ∇ is **metric compatible** if

$$\nabla_{\boldsymbol{u}}\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{g}(\nabla_{\boldsymbol{u}}\boldsymbol{v}, \boldsymbol{w}) + \boldsymbol{g}(\boldsymbol{v}, \nabla_{\boldsymbol{u}}\boldsymbol{w}).$$

In terms of components, the requirement of metric compatibility is equivalent to $\nabla_k g_{ij} = 0$. The same holds also for the components of the inverse metric tensor: $\nabla_k g^{ij} = 0$. Thus the covariant derivative with respect to a metric compatible connection commutes with raising and lowering indices. An another important property of metric compatible connections is that the length of a tangent vector remains constant under parallel transport along a geodesic [5].

The covariant derivative also allows us to define a quantity called *torsion*, that is obtained from $\mathcal{T}(\boldsymbol{v},\boldsymbol{w}) \equiv \nabla_{\boldsymbol{v}}\boldsymbol{w} - \nabla_{\boldsymbol{w}}\boldsymbol{v} - [\boldsymbol{v},\boldsymbol{w}]$. The second building block needed for the construction of a unique connection is to require the connection to be torsion-free by setting $\mathcal{T}(\boldsymbol{v},\boldsymbol{w}) = \boldsymbol{0}$. In components this is equivalent to demanding that the Christoffel symbols must be symmetric in their lower indices such that $\mathcal{T}^{\lambda}_{\mu\nu} \equiv \Gamma^{\lambda}_{\mu\nu} - \Gamma^{\lambda}_{\nu\mu} = 0$. The unique torsion-free and metric compatible connection constructed this way is known as the *Levi-Civita connection* [2]. Its Christoffel symbols are given by

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}).$$

Now we are in a position to define some tensor quantities to describe the curvature of a Riemannian manifold. In the definition below, we make use of the *commutator* $[\boldsymbol{u},\boldsymbol{v}] \equiv \boldsymbol{uv} - \boldsymbol{vu}$ of two vector fields $\boldsymbol{u}$ and $\boldsymbol{v}$.

Definition 6.3. Let $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ be vector fields on a Riemannian manifold. The **Riemann endomorphism** or the *curvature operator* is given by

$$\mathcal{R}(\boldsymbol{u},\boldsymbol{v})\boldsymbol{w} \equiv [\nabla_{\boldsymbol{u}},\nabla_{\boldsymbol{v}}]\boldsymbol{w} - \nabla_{[\boldsymbol{u},\boldsymbol{v}]}\boldsymbol{w}. \tag{6.2}$$

Acting on the endomorphism with the flat operation gives us the **Riemann curvature tensor**[4] $\boldsymbol{R}$

$$\mathcal{R}^\flat \equiv \boldsymbol{R} = R_{ijkl}\, dx^i \otimes dx^j \otimes dx^k \otimes dx^l,$$

which, with the conventions of [3], acts on four vector fields $\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{x},\mathcal{R}(\boldsymbol{u},\boldsymbol{v})\boldsymbol{y}).$$

This is to say that the index to be lowered by the flat operation is the first one, such that we may write the action of the Riemann endomorphism on some basis vectors $\partial_i, \partial_j, \partial_k$ as

$$\mathcal{R}(\partial_i,\partial_j)\partial_k = R^l{}_{ijk}\partial_l.$$

The above equation displays why the Riemann endomorphism is sometimes called the curvature operator. Its usefulness comes from the *flatness criterion*, demanding it to be zero for Euclidean spaces [2]. Hence the Riemann endomorphism, and the tensor just as well, contain information on how the manifold differs from the usual flat Euclidean space. That is to say, how *curved* it is. It follows from the above definition that after raising one of the indices, the components of the Riemann tensor are given by

$$R^l{}_{ijk} = \partial_j \Gamma^l_{ki} - \partial_k \Gamma^l_{ji} + \Gamma^l_{jm}\Gamma^m_{ki} - \Gamma^l_{km}\Gamma^m_{ji}.$$

Now let $\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}$ be vector fields on a Riemannian manifold. The Riemann tensor has the symmetries [2, 3]

$$\begin{aligned}
\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}) &= -\boldsymbol{R}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) \\
\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}) &= -\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{v},\boldsymbol{u}) \\
\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}) &= \boldsymbol{R}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \\
\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}) &+ \boldsymbol{R}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y},\boldsymbol{v}) + \boldsymbol{R}(\boldsymbol{y},\boldsymbol{u},\boldsymbol{x},\boldsymbol{v}) = 0.
\end{aligned}$$

[4] Often called just the Riemann tensor for brevity.

By contracting the first and third indices of the Riemann tensor components we get the components of the *Ricci tensor*

$$R_{ij} \equiv R^l{}_{ilj}$$

and tracing the components of the Ricci tensor results in the *Ricci scalar*

$$R \equiv R^i{}_i.$$

For our later discussion on the variations of arclength, we still need to introduce a map that takes a tangent vector living on the tangent space of a manifold to the corresponding geodesic on the manifold itself and proceeds along it for unit time starting from the point x. To do this, we use the idea that at each point of a Riemannian manifold there is a unique maximal geodesic corresponding to each vector in the tangent space at that point [2].

Definition 6.4. Let M be a manifold and χ a subset of TM containing all the vectors $\boldsymbol{v} \in TM$ for which the unique maximal geodesic $\gamma_{\boldsymbol{v}}$ is defined on an interval that contains $[0, 1]$. Then the **exponential map** is defined as $\exp(\boldsymbol{v}) \equiv \gamma_{\boldsymbol{v}}(1)$. By restricting the domain χ to $\chi_x \equiv \chi \cap T_x M$, we get the restricted exponential map $\exp_x$.

Next, let us study an important property of the $\exp_x$ defined above. Suppose the coordinates on the geodesic $\gamma_{\boldsymbol{v}}$ obtained with some value $\tau \in [0, 1]$ of the curve parameter are given by $\gamma^i(\tau)$. Now take these coordinates to be obtained from the components of the tangent vector $\boldsymbol{v}$ by the relation

$$\gamma^i(\tau) = \tau v^i$$

Such coordinates are referred to as the *Riemann normal coordinates* or *local geodesic coordinates*. This is because the second derivative gives

$$\frac{d^2\gamma^i}{d\tau^2} = 0$$

and comparison with the geodesic equation yields that we must have

$$\Gamma^i_{jk}\big|_x = 0.$$

Here $\big|_x$ stands for evaluation at point $x \in M$. This is to say that we have a local coordinate system where the connection coefficients vanish. When calculating in terms of tensor components, the existence of such a coordinate system tends to simplify most computations radically. Furthermore, notice that there is no loss of generality in deriving tensor relations in this coordinate system due to the coordinate-free nature of tensors [6, 7].

7. Pseudo-Riemannian metrics and Lorentzian geometry

By choosing to relax some or one of the requirements that define a Riemannian metric, we arrive at more general metrics. This happens in the sense that every such metric is not necessarily Riemannian, but the more strict defining requirements of Riemannian metrics still satisfy the remaining requirements, making Riemannian metrics instances of the more general metrics. Of special interest to applications in theoretical and mathematical physics are **pseudo-Riemannian metrics**, for which we drop the requirement of positive definiteness. The requirement that a manifold M be locally Euclidean is then replaced by demanding the existence of a local coordinate frame where the metric tensor can be represented as a diagonal matrix so that

$$g_{\mu\nu}dx^{\mu}dx^{\nu} = -(dx^0)^2 - ... - (dx^{r-1})^2 + (dx^r)^2 + ... + (dx^{\dim(M)-1})^2.$$

On pseudo-Riemannian manifolds our coordinate labels start from 0. The amount of negative terms, r, is the *index* of the metric. *Sylvester's law of inertia* states that the index is the maximum dimension of any subspace of a tangent space on a pseudo-Riemannian manifold for which the diagonalized metric tensor is negative definite. The index is thus independent of the choice of basis [2]. When considering pseudo-Riemannian manifolds, we shall use greek letters as the sub- and superscripts of tensor components. Roman letters are reserved for Riemannian manifolds and submanifolds (see section 8 for a brief discussion of pseudo-Riemannian submanifolds). Now it is noteworthy that despite we referred to Riemannian manifolds in sections 5 and 6, the same concepts exist equally in pseudo-Riemannian geometry. The necessary modifications are to convert roman sub- and superscripts into greek ones and to account for the signature of the inner product when considering the length functional, as is eventually done in section 9.

In classical mechanics we are usually concerned with problems in which the space is a Riemannian manifold, for instance $\mathbb{R}^n$. The path of a particle can be parametrized by the time t, an absolute parameter measured to be the same by all observers everywhere on the manifold. Nevertheless, to see the physical significance of the properties of pseudo-Riemannian manifolds, let us recall some central ideas in the special theory of relativity. First, any observer O in some inertial frame considers oneself to be at rest, whilst everything that is not in the same frame appears to be in motion relative to O. Second, there is a finite maximum signal speed which is measured to be the same natural constant c by all observers. What is more, we must demand consistency between coordinate transformations. These remarks inevitably enforce us to conclude that any measured lengths and time-intervals are entirely frame-dependent. For this to make any sense, the role of time as an absolute parameter must be rejected. The role of the constant c is then to make the units of time and space agree so that the idea of a Riemannian manifold as the fundamental model of the Universe is replaced by *spacetime*, a pseudo-Riemannian manifold of index one, also referred to as a *Lorentzian manifold*. The geometry of such manifolds then naturally goes by the name of Lorentzian geometry. When referring to relativity, we shall assume that spacetime is a four-dimensional Lorentzian manifold with three spatial dimensions. The coordinate time-related items of a vector $\boldsymbol{x}$ are placed into the zeroth component x^0. Spatial components are denoted by x^i, so that altogether we have x^{μ} with $\mu \in \{0, 1, 2, 3\}$. A point in spacetime is called an *event*.

We must now enable ourselves to work with the tangent spaces of Lorentzian manifolds. As with the vector spaces of Riemannian manifolds, the inner product of two vectors in the tangent space of a Lorentzian manifold is again given by the metric tensor. One ought to be aware that, contrary to the usual definitions of an inner product, we must now allow the inner product of a non-zero vector with itself to yield also negative values or zero. This motivates us to assign the vectors in a tangent space of a Lorentzian manifold into three types of subspaces.

Definition 7.1. Let M be a Lorentzian manifold with a metric $\boldsymbol{g}$ and $x \in M$ a point. A subspace $W \subset T_xM$ is called
(i) **timelike** if $\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{w}) < 0$ for all $\boldsymbol{w} \in W$.
(ii) **spacelike** if $\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{w}) > 0$ for all $\boldsymbol{w} \in W$.
(iii) **lightlike** or **null** if $\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{w}) = 0$ for all $\boldsymbol{w} \in W$.
In this naming scheme, the character of a vector $\boldsymbol{w} \in W$ is the same as that of W.

Paths in spacetime can be classified analogously, for instance such that the tangent vector of a timelike path $\gamma(\tau)$ is timelike everywhere along the path. The set of all null vectors at a a certain event forms two *lightcones*, separating the timelike and spacelike vectors in T_xM. One of the cones contains all timelike and null vectors pointing to the event's past, whereas the other one then contains those pointing to the future. Intrinsically, we may not have a way to classify which cone is which; to make such a choice is to pick a *time-orientation* on the manifold [4].

The above concepts enable us to discuss the possible causal relations between events as well as introducing some peculiarities of Lorentzian geometry. For illustration, see Figure 1.

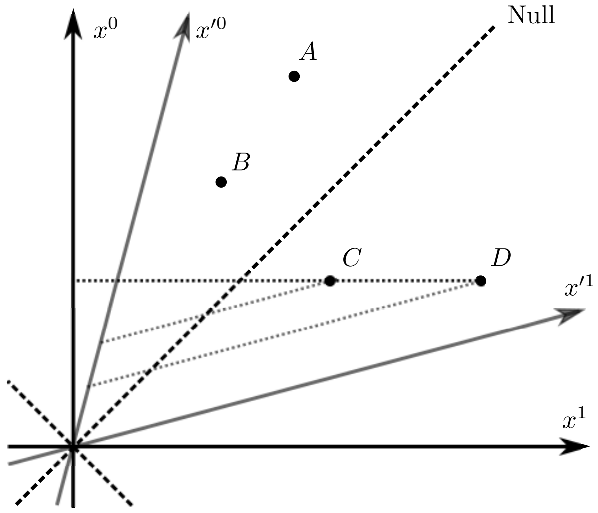

FIGURE 1. There is no ambiguity on the possible causal relation between the two timelike separated events A and B. However, observers in the unprimed coordinate system would calculate that the spacelike separated C and D occurred simultaneously with respect to their coordinate time, whereas any observer in the primed system would arrive at a different answer.

To state causality in more rigorous terms, let M be a time-oriented Lorentzian manifold and $x \in M$ an event with some event $y \in M$ in its future lightcone. The usual notation for saying that there is a timelike path between the events is $x \ll y$. If at least a null path between the events exists, we write $x < y$. $x \leq y$ stands for $x < y$ or $x = y$. We then have the following definition.

Definition 7.2. Let M be a Lorentzian manifold with a subset $A \subset M$. The causal future $I^+(A)$, chronological future $J^+(A)$, causal past $I^-(A)$ and chronological past $J^-(A)$ of A are defined as [4]

$$I^+(A) \equiv \{y \in M | \text{ There is a } x \in A \text{ such that } x \ll y\},$$
$$J^+(A) \equiv \{y \in M | \text{ There is a } x \in A \text{ such that } x < y\},$$
$$I^-(A) \equiv \{y \in M | \text{ There is a } x \in A \text{ such that } x \gg y\},$$
$$J^-(A) \equiv \{y \in M | \text{ There is a } x \in A \text{ such that } x > y\}.$$

The classically intuitive idea of the spatial Universe as a Riemannian manifold is replaced in relativity by introducing *spacelike hypersurfaces*. Of special interest is the classification of **Cauchy hypersurfaces** as any subset S of a pseudo-Riemannian manifold M that is pierced exactly once by all inextensible timelike curves on M. These are illustrated in Figure 2. Further, any set that is met at most once by any timelike inextensible curve is dubbed *achronal* [4]. Here the demand of inextendibility is to say that we limit our consideration to curves that are not merely segments of some other curve but, rather, could not be extended any further.

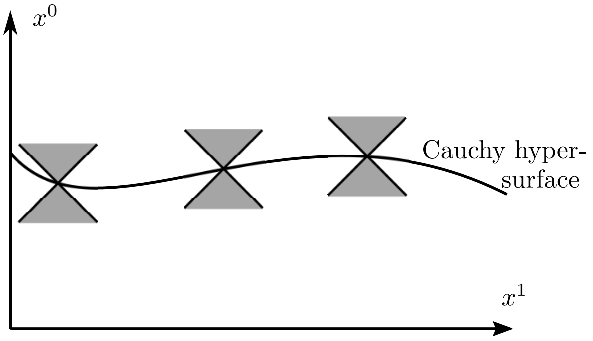

FIGURE 2. An example of a Cauchy hypersurface on a two-dimensional pseudo-Riemannian manifold with light cones attached to some points on the surface. Notice that the Cauchy hypersurface may never re-enter any light cone once starting elsewhere and elsewhen on it.

Definition 7.3. Let A be an achronal subset of a pseudo-Riemannian manifold M. The future **Cauchy development** $D^+(A)$ of A is the set

$$D^+(A) = \{x \in M | \text{all past inextensible causal curves through } x \text{ pierce } A\}.$$

The past Cauchy development $D^-(A)$ is defined analogously using future inextensible causal curves.

We should enable ourselves to discuss the regions of a Lorentzian manifold M over which any event cannot be determined by some achronal subset $A \subset M$. For instance, if an observer starting at A encounters an insuperable obstacle given by some other subset $B \subset M$, the future Cauchy development of B will contain points over which A has no say, but which may still affect A's future. Such cases require the notion of *Cauchy horizons* as the border between what can and what cannot be determined by A. An illustration of the rise and usefulness of some of these sets is given in Figure 3.

Definition 7.4. Let M be a Lorentzian manifold with an achronal subset $A \subset M$. The **future Cauchy horizon** of A is defined as [4]

$$H^+(A) \equiv \overline{D^+(A)} \backslash I^-(D^+(A)),$$

where $\overline{D^+(A)}$ denotes the *closure* of $D^+(A)$, the smallest closed set containing $D^+(A)$. The past Cauchy horizon is defined as

$$H^-(A) \equiv \overline{D^-(A)} \backslash I^+(D^-(A)).$$

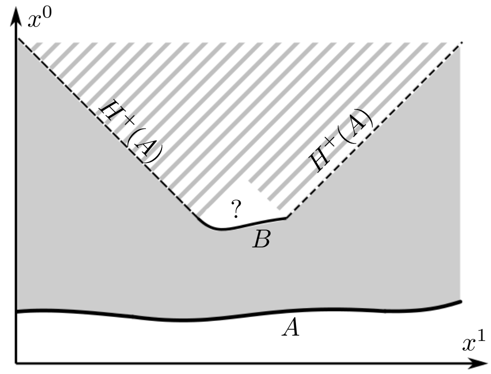

FIGURE 3. An illustration of the difference between the Cauchy development $D^+(A)$ (gray) and the chronological future $J^+(A)$ (gray+striped), when A is an achronal set depicted by the thick curve and we are given the set of events B via which any timelike path starting at A may not trespass. This gives rise to the Cauchy horizons $H^+(A)$ drawn as dashed lines. The area marked by the question mark contains all unforeseeable events that may affect the future of A.

Our parametrizations of some path in spacetime – a *worldline* – is motivated by the idea that the time measured by the observer traveling on the path should always be measured in the same way by the said observer. This may not necessarily be the coordinate time x^0, but it should be tied to a measurement of the constant c, which the observer may always perform and which ought to give the same result at any event. This, loosely speaking, results in the time measured by a personal clock carried by the observer. We call this the *proper time* τ and shall in general take it to be the curve parameter of timelike paths. More precisely, τ is given by the squared line-element via

$$c^2 d\tau^2 \equiv -ds^2.$$

Rationally, the value of τ may only increase[5] for any physical observer who is to measure the maximal signal speed as c. No matter how fast the observer goes, c will never become 0 relative to the observer. With the four dimensional pseudo-Riemannian index one metrics of relativity, if we parametrize timelike curves by proper time, geodesics become the curves of maximum proper time [6].

The spacetime of special relativity is known as the *Minkowski spacetime*. With local coordinates $\{cx^0, x^1, x^2, x^3\}$, the squared line-element becomes

$$ds^2 = -(cdx^0)^2 + \sum_{i=1}^{3}(dx^i)^2.$$

From here onwards we shall use *natural units* and set $c = 1$. Adopting the conventions of [3], the components of the pseudo-Riemannian metric of Minkowski spacetime can be represented as

[5]Without treading too far into the domain of metaphysics, that is. Should τ ever start to decrease, the direction of all experienced time would be reversed. This would consequently erase all historical records of such a bizarre phenomenon, also including any observer's memory. Then it is no more a question of physics whether or not such a thing ever really happened.

$g_{\mu\nu} \doteq \text{diag}(-1, 1, 1, 1)$. Now recall that any change $\Delta\tau$ obtained from $d\tau^2$ must always be positive. It is in fact the one differing sign in the metric that provides us with a way of ensuring causal connections between two timelike separated events to be unchanged in coordinate transforms.

Instead of being locally Euclidean in the fashion of Riemannian manifolds, a general 4 dimensional pseudo-Riemannian spacetime of index 1 is locally Minkowski. Therefore, even in the curved spacetimes of general relativity, the laws of physics always reduce to those of special relativity in local frames. This is the *Einstein equivalence principle* [7, 6]. The main focus of the latter Lorentzian part of this text shall be on Hawking's theorem, which is concerned with the *global* properties of a spacetime and not only the physics of its local frames. We shall return to the theorem after first discussing the analogous Myers's theorem in Riemannian geometry, but at this point it is instructive to introduce some central ideas of the general theory of relativity. We do so by means of introducing some local equations, which however are tensor relations and hence coordinate-invariant.

In general relativity we build on the principles of special relativity, but wish to include a geometric description of gravity. The gravitational force of classical mechanics is replaced by stating that the presence of energy[6] causes spacetime curvature. The phenomenon of objects "falling" under gravity then arises from how their paths through spacetime are affected by its curvature. Notice that a massive object always moves forward in *spacetime* even if it is at rest in *space*, since its worldline parameter is proceeding. Let us consider a particle at constant spatial coordinates in some frame. In Minkowski spacetime, its worldline is a straight line parallel to the local coordinate time axis. For the sake of illustration, let us say the worldline is parametrized by the proper time. Now introduce a source of gravity, causing spacetime to curve. The worldline – being embedded in the spacetime – then bends accordingly, which leads to an evolving spatial displacement as proper time elapses.

In continuum mechanics, where matter is described as a continuous medium, we have for a non-viscous fluid that

$$-\partial_0\rho + \partial_i p^i = 0$$

ergo whenever the pressure p^i changes, there must be a flow of matter changing the density ρ as a function of time. This is known as the *continuity equation.* Assuming the energy-momentum tensor $\boldsymbol{T}$ to be that of a *perfect fluid* with the components $T^{\mu\nu} \doteq \text{diag}(\rho, p, p, p)$ in the fluid rest frame, we may write the continuity equation in Minkowski spacetime as $\partial_\mu T^{\mu\nu} = 0$. However, we need to be able to expand our attention to more general spacetimes and fluid models. A somewhat more fundamental requirement, reducing to the continuity equation in the above case, is that $\boldsymbol{T}$ needs to be *four-divergence-free.* The usual prescription to replace a physical law that holds in Minkowski spacetime with one that holds for more general Lorentzian manifolds is to change all partial derivatives into covariant derivatives [3, 6, 7]. The idea is that if a physical law is to be applicable everywhere on the manifold, it must be possible to state it as a tensor relation. However, the partial derivatives of tensor components are not the components of a new tensor. This is remedied by the introduction of the covariant derivative, which reduces to the partial derivative on flat spacetimes, as should. Thus

$$\nabla_\mu T^{\mu\nu} = 0.$$

Now if the four-divergence-free energy-momentum tensor is to contain all information regarding the source(s) of gravitation, it needs to be connected to spacetime curvature. Since the energy-momentum tensor is divergence-free and has rank 2, these must also be the properties of some

[6] Recall that energy and mass are equivalent in relativity.

tensor $\boldsymbol{G} = G^{\mu\nu}\partial_\mu \otimes \partial_\nu$ that is to be constructed out of the curvature tensors presented in previous sections. Thus

$$G^{\mu\nu} = \kappa T^{\mu\nu}, \tag{7.1}$$

where κ is a natural constant depending our choice of units. Since the rank is 2, our possibilities are limited to some combination of the Ricci tensor and the Ricci scalar. Of course we could, as is done in the case of the cosmological constant, add also another term $\Lambda g_{\mu\nu}$ with some scalar Λ to the right-hand side. Furthermore, there are no terms including products of the Riemann tensor since, in the Levi-Civita connection, $R^{\mu}{}_{\nu\alpha\beta}$ is linear in the second partial derivatives of the metric tensor components. This is sufficient to explain the Newtonian limit of gravity described by Poisson's equation [7]. It nevertheless turns out [3, 6, 7] that the combination suitable for much interesting geometry and physics is

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu}. \tag{7.2}$$

Equation (7.1) is a set of differential equations known as the *Einstein field equations.* Their importance lies in the fact that, assuming the validity of general relativity, all physically interesting spacetimes are those that satisfy the equations (7.1). This is to say that the metric leading to a solution of the Einstein equations is the description of a spacetime, whose energy content is given by the energy-momentum tensor of some model for a cosmic fluid.

8. Properties of pseudo-Riemannian submanifolds

We denote that a manifold $\tilde{M}$ is a *submanifold* of some other ambient manifold M by $\tilde{M} \subset M$. The submanifold is essentially any set of points within the ambient manifold that qualifies as a manifold on its own. A straightforward illustration is given by the Minkowksi spacetime. Its spatial part is described by $\mathbb{R}^3$, which is a Riemannian manifold. Further, we may consider the two-dimensional sphere to be a submanifold of $\mathbb{R}^3$. Then, as anything that is Riemannian is also pseudo-Riemannian, the sphere and $\mathbb{R}^3$ are both pseudo-Riemannian submanifolds of the Minkowski spacetime. For brevity, we shall refer to any submanifolds of a pseudo-Riemannian or a Lorentzian manifold as either pseudo-Riemannian or Lorentzian submanifolds, regardless if the submanifold itself is either Riemannian or pseudo-Riemannian. The study of submanifolds offers a rich topic in both Riemannian and pseudo-Riemannian geometry. Anyhow, for the scope of this text, we shall restrict our attention to the concepts needed for discussing Hawking's theorem in section 13.

Let M be a Riemannian or pseudo-Riemannian manifold with the Levi-Civita connection ∇ and a submanifold $\tilde{M}$. The first property of submanifolds we are concerned with is that sections of

$$TM|_{\tilde{M}} \equiv \coprod_{x\in\tilde{M}} T_xM$$

can be decomposed into parts tangential and orthogonal to $\tilde{M}$. Then we may express the tangent space of M at a point x on the submanifold $\tilde{M}$ as the sum

$$T_xM = (T_x\tilde{M})_\perp + T_x\tilde{M},$$

where $(T_x\tilde{M})_\perp$ stands for the space of all vectors in T_xM that do not belong to $T_x\tilde{M}$. We may now write the covariant derivative as

$$\nabla_{\boldsymbol{u}}\boldsymbol{v} = (\nabla_{\boldsymbol{u}}\boldsymbol{v})_\perp + (\nabla_{\boldsymbol{u}}\boldsymbol{v})_\|.$$

The tangential component $(\nabla_{\boldsymbol{u}} \boldsymbol{v})_{\parallel}$ turns out to be the covariant derivative on $\tilde{M}$. The orthogonal term motivates the introduction of the **second fundamental form** as [2]

$$\boldsymbol{I\!I}(\boldsymbol{u}, \boldsymbol{v}) \equiv (\nabla_{\boldsymbol{u}} \boldsymbol{v})_{\perp}.$$

To avoid confusion, one should pay attention to what is meant by *orthogonality* in the case of pseudo-Riemannian manifolds: a vector belonging to the tangent space of M at a point $x \in \tilde{M} \subset M$ may be divided into a sum of some vector that belongs to $T_x\tilde{M}$ and a vector that does not. We say that the latter is orthogonal to $\tilde{M}$.

We obtain the mean normal curvature $\boldsymbol{h}$ of a pseudo-Riemannian submanifold from the second fundamental form $\boldsymbol{I\!I}$ as follows. Take M to be a pseudo-Riemannian manifold of dimension n with an orthonormal basis $\{\boldsymbol{e}_\mu\}$ for T_xM. Suppose the submanifold $\tilde{M}$ is a spacelike hypersurface of dimension $n-1$ ergo the inner product defined by the metric on the submanifold is positive-definite. Notice that we shall denote the metric tensor on both M and $\tilde{M}$ by $\boldsymbol{g}$, since the metric on $\tilde{M}$ is merely a restriction of that on M with one or more dimensions dropped. Notice further that we can for instance take the inner product of a vector in $T_x\tilde{M}$ and a vector in T_xM by granting the vector in $T_x\tilde{M}$ the missing dimensions and writing it as a vector in T_xM. We then have [4]

$$\boldsymbol{h}_x = \frac{1}{n-1} \sum_{i=1}^{n-1} \boldsymbol{I\!I}(\boldsymbol{e}_i, \boldsymbol{e}_i).$$

This enables us to define a *shape operator* and *convergence* on pseudo-Riemannian submanifolds [4].

Definition 8.1. Let M be a pseudo-Riemannian manifold with a pseudo-Riemannian hypersurface $\tilde{M} \subset M$. Suppose $\tilde{M}$ has a future pointing unit normal vector field $\boldsymbol{u}$. The **shape operator** of the hypersurface is a $(1,1)$-tensor field $\boldsymbol{S_u}$ on $\tilde{M}$ defined by the relation

$$\boldsymbol{g}(\boldsymbol{S}_u(\boldsymbol{v}), \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{I\!I}(\boldsymbol{v}, \boldsymbol{w}), \boldsymbol{u}),$$

where $\boldsymbol{v}$ and $\boldsymbol{w}$ are smooth vector fields on $\tilde{M}$ [4].

Definition 8.2. Let M be a pseudo-Riemannian manifold of dimension n with a hypersurface of dimension $n-1$ as a submanifold, on which $\boldsymbol{u}$ is a future pointing unit normal vector field. The **convergence** of the hypersurface is defined as

$$\boldsymbol{k}^*(\boldsymbol{u}) = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{h}) = (\dim\tilde{M})^{-1} \mathrm{Tr} \boldsymbol{S_u},$$

where $\boldsymbol{h}$ is the mean normal curvature of $\tilde{M}$ and $\boldsymbol{S_u}$ is the shape operator of the hypersurface [4].

9. First variation of arclength

The central idea of calculus of variations is to find an equation that extremizes the value of the integral of a *functional*, a function that takes other functions as input arguments.

Definition 9.1. Let $\gamma_0(\tau)$ be a smooth curve on a pseudo-Riemannian manifold M. We call γ_0 the main curve of variation. A **variation** of γ_0 is the smooth map from a set of points $[\tau_0, \tau_1] \times [\alpha_0, \alpha_1] \in \mathbb{R}^2$ (the parameter space) to M such that for each line with a constant value of α in the parameter space we have an α-curve $\gamma_\alpha(\tau)$ on M. Correspondingly, the lines of constant τ value are mapped to τ-curves $\gamma_\tau(\alpha)$ [2, 5]

The usual treatment of the problems in calculus of variations, as found in some physics textbooks, involves adding infinitesimal offsets to the arguments of the functional to be studied, Taylor expanding it and demanding that the first order term vanishes for extremizing solutions. Here we however perform a calculation inspired by [2, 5], based on the idea that the extrema of a function are found by setting its first derivative to zero. Additionally, we make sure to proceed in a manner suitable for both Riemannian and Lorentzian manifolds. Notice that the approach here does not enforce us to demand the variation to be *proper* i.e. the variations need not vanish at the end points of the curve to be varied. The method displayed here neatly showcases the power of all the formalism developed so far. Throughout sections 9 – 11, we shall use the quantities given in the above definition and illustrated in Figure 4.

Definition 9.2. Let $f : M \to \mathbb{R}$ be a differentiable functional such that

$$f(\gamma + \xi) - f(\gamma) = f'(\xi) + \mathcal{O}(\xi^2)$$

for all curves[7] γ and variation curves ξ on M such that $\gamma + \xi$ is also a curve on M and f' is a linear functional. $\mathcal{O}(\xi^2)$ denotes possible nonlinear terms. We say that the differential f' is the **first variation** of f.

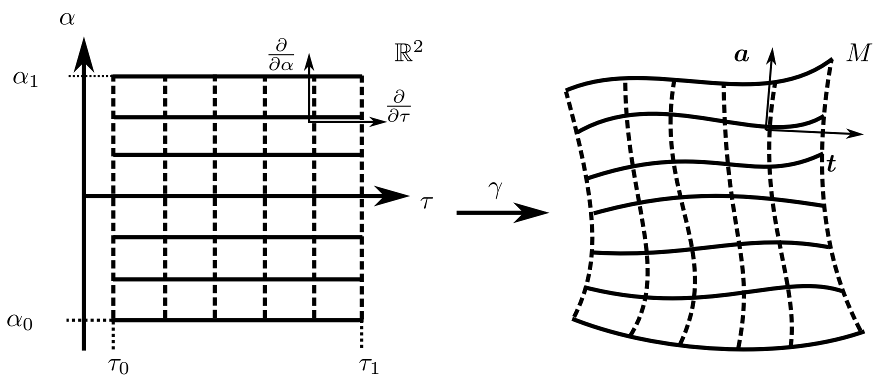

FIGURE 4. γ maps the parameter space $[\alpha_0, \alpha_1] \times [\tau_0, \tau_1] \in \mathbb{R}^2$ to the manifold M. The solid lines become what are known as a-curves and the dashed lines t-curves. The vectors $\boldsymbol{t}, \boldsymbol{a} \in TM$ arise from pushing forward the basis vectors $\partial/\partial\tau$ and $\partial/\partial\alpha$, respectively.

Theorem 9.3. *Let M be a Riemannian or a pseudo-Riemannian manifold with a metric compatible and torsion-free connection. Denote the signature of the inner product by* $\operatorname{sgn}(\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha)) \equiv \varepsilon$. *The first variation of arclength s is given by*

$$s' = \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}\right) d\tau, \tag{9.1}$$

without the necessity to demand a proper ergo a fixed end-point variation.

[7]The curve parameters have been omitted for brevity.

Proof. In order to handle both the Riemannian and timelike Lorentzian case in the same proof, we replace the arclength by a proper time interval for timelike Lorentzian curves, but stick to denoting both by s for simplicity. This also cancels the negative sign in the square root of the length functional for timelike curves, which shall be the most interesting ones for our later applications. Eq. (5.2) then becomes

$$s(\alpha) = \int_{\tau_0}^{\tau_1} \sqrt{\varepsilon \boldsymbol{g}\left(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha\right)} d\tau.$$

We find the first variation of arclength by differentiating this with respect to the parameter α. This corresponds to acting on the integral with $\boldsymbol{a}$, the pushforward of the parameter space basis vector $\partial/\partial\alpha$. As $\boldsymbol{a}$ is τ-independent here, we may move it inside the integral.

$$\begin{aligned}
s'(\alpha) &= \int_{\tau_0}^{\tau_1} \boldsymbol{a}\sqrt{\varepsilon \boldsymbol{g}\left(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha\right)} d\tau \\
&= \int_{\tau_0}^{\tau_1} \frac{1}{2}\left(\varepsilon \boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha)\right)^{-1/2} \varepsilon \boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) d\tau.
\end{aligned}$$

Using a local coordinate basis $\{\partial_\mu\}$ on the manifold and writing the components of $\boldsymbol{t}_\alpha$ as $\dot{x}^\mu$, we notice that

$$\begin{aligned}
\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) &= a^\rho \partial_\rho (g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu) \\
&= a^\rho \left((\partial_\rho \dot{x}^\mu)\dot{x}_\mu + \dot{x}^\mu \partial_\rho \dot{x}_\mu\right) \\
&= a^\rho \left((\nabla_\rho \dot{x}^\mu - \Gamma^\mu_{\rho\nu}\dot{x}^\nu)\dot{x}_\mu + \dot{x}^\mu(\nabla_\rho \dot{x}_\mu + \Gamma^\nu_{\rho\mu}\dot{x}_\nu)\right).
\end{aligned}$$

As the dummy indices μ and ν in the last term are interchangeable, the two terms with the Christoffel symbols cancel. We may then proceed with

$$\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) = a^\rho \left((\nabla_\rho \dot{x}^\mu)\dot{x}_\mu + \dot{x}^\mu(\nabla_\rho \dot{x}_\mu)\right).$$

Due to metric compatibility, this is equivalent to

$$\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) = \boldsymbol{g}\left(\nabla_{\boldsymbol{a}}\boldsymbol{t}, \boldsymbol{t}\right) + \boldsymbol{g}\left(\boldsymbol{t}, \nabla_{\boldsymbol{a}}\boldsymbol{t}\right).$$

The symmetry of the metric tensor lets us combine the two terms, resulting in

$$\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) = 2\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha)).$$

Notice that the usage of proper time instead of arclength in the Lorentzian case also demands us to account for the signature in the norm of the tangent vector: $\left(\varepsilon \boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha)\right)^{1/2} = |\boldsymbol{t}_\alpha|$. Thus we may write

$$\begin{aligned}
s' &= \varepsilon \int_{\tau_0}^{\tau_1} |\boldsymbol{t}_\alpha|^{-1} \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha) d\tau \\
&= \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha\right) d\tau.
\end{aligned}$$

For a torsion-free connection we have $\mathcal{T}(\boldsymbol{a}, \boldsymbol{t}_\alpha) = \boldsymbol{0}$, which can be written as

$$\begin{aligned}
\nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha - \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a} - [\boldsymbol{a}, \boldsymbol{t}_\alpha] &= \boldsymbol{0} \\
\nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha &= \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a},
\end{aligned}$$

where $[\boldsymbol{a}, \boldsymbol{t}_\alpha] = \boldsymbol{0}$ follows from $[\partial_\alpha, \partial_\tau] = \boldsymbol{0}$. Hence, the first variation becomes

$$s' = \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}\right) d\tau.$$

□

It is also instructive to process the first variation of arclength a bit further to see how the usual fixed end-point result comes out and what are the resulting geometric objects. Utilizing the product rule and noticing that $\nabla_{\boldsymbol{t}_\alpha}$ operating on a scalar is $\partial/\partial\tau$, we may write the above result for the first variation in the alternative form

$$\begin{aligned} s' &= \varepsilon \int_{\tau_0}^{\tau_1} \frac{\partial}{\partial\tau}\boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \boldsymbol{a}\right) - \boldsymbol{g}\left(\nabla_{\boldsymbol{t}_\alpha}\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \boldsymbol{a}\right) d\tau \\ &= \varepsilon\boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \boldsymbol{a}\right)\Bigg|_{\tau_0}^{\tau_1} - \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{g}\left(\nabla_{\boldsymbol{t}_\alpha}\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \boldsymbol{a}\right) d\tau. \end{aligned}$$

If we demand that the first variation s' vanishes for arbitrary transversal vector fields $\boldsymbol{a}$ and demand $|\boldsymbol{t}_\alpha| = 1$ for the main curve, the only solution is that $\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{t}_\alpha = \boldsymbol{0}$. As $\boldsymbol{t}_\alpha$ is the tangent vector of the curve solving our variational problem, we may write this as $D_\tau\boldsymbol{t}_\alpha = \boldsymbol{0}$. This is nothing but the autoparallelity requirement, telling us that the extremizing curves are geodesics.

10. The Jacobi equation and Jacobi fields

Let us continue with the vector fields $\boldsymbol{t}$ and $\boldsymbol{a}$ of section 9 such that $\boldsymbol{t}$ is a tangent vector field and $\boldsymbol{a}$ a transversal field along a geodesic γ. As was stated in section 6, this means that the autoparallelity requirement

$$D_\tau\boldsymbol{t} = \boldsymbol{0} \tag{10.1}$$

must be satisfied. Taking the covariant derivative of Eq. (10.1) with respect to the transversal field yields

$$D_\alpha D_\tau\boldsymbol{t} = \boldsymbol{0}. \tag{10.2}$$

Since $[\boldsymbol{a}, \boldsymbol{t}] = \boldsymbol{0}$ as in section 9, the last term of the Riemann endomorphism (6.2) vanishes and we may rewrite it using the covariant derivatives along γ as

$$\mathcal{R}(\boldsymbol{a}, \boldsymbol{t})\,\boldsymbol{t} = D_\alpha D_\tau\boldsymbol{t} - D_\tau D_\alpha\boldsymbol{t}. \tag{10.3}$$

Inserting Eq. (10.2) into Eq. (10.3) and rearranging the terms, we get

$$\mathcal{R}(\boldsymbol{a}, \boldsymbol{t})\,\boldsymbol{t} + D_\tau D_\alpha\boldsymbol{t} = \boldsymbol{0}.$$

Similarly to section 9, we have for zero torsion that $D_\tau\boldsymbol{a} = D_\alpha\boldsymbol{t}$. Utilizing this gives us the **Jacobi equation**

$$D_\tau^2\boldsymbol{a} + \mathcal{R}(\boldsymbol{a}, \boldsymbol{t})\,\boldsymbol{t} = \boldsymbol{0}. \tag{10.4}$$

A *Jacobi field* is a vector field $\boldsymbol{a}$ satisfying the above [2]. An important application of Jacobi fields is that they let us define *conjugate points* along a geodesic.

Definition 10.1. Let p, q be two points along a geodesic γ. q is a **conjugate point** of p if there is a Jacobi field along γ that is not identically zero but vanishes at p and q.

For cases where the other end of the geodesic is not on a specific point but somewhere on a pseudo-Riemannian submanifold, we generalize conjugate points by introducing *focal points* [4]. They have the interesting feature that the exponential map is singular at them. For further discussion on this see e.g. [8, 9].

Definition 10.2. Let M be a pseudo-Riemannian manifold with a submanifold $\tilde{M} \subset M$. Suppose $\gamma(\tau)$ is a geodesic on M such that $\gamma(0)$ is a point on $\tilde{M}$ and the tangent vector of γ at $\tau = 0$ is $\boldsymbol{t}(0) \in (T_{\gamma(0)}\tilde{M})_\perp$. We say that γ is orthogonal to $\tilde{M}$. A point $p \neq \gamma(0)$ along γ is then a **focal point** of $\tilde{M}$ if there is a Jacobi field $\boldsymbol{J}(\tau)$ along γ satisfying the conditions

(i) $\boldsymbol{J}(0) \in T_{\gamma(0)}\tilde{M}$,
(ii) $(D_\tau \boldsymbol{J}|_{\tau=0})_\parallel = (\nabla_{\boldsymbol{J}(0)}\boldsymbol{t}(0))_\parallel$,
(iii) $\boldsymbol{J}$ is not identically zero but vanishes at p.

A common example of a Jacobi field given in textbooks on general relativity is a vector field whose magnitude is related to the measured spatial separation of particles traveling along nearby geodesics. For such a Jacobi field, the Jacobi equation (10.4) can be considered to show that the presence of curvature as described by $\mathcal{R}$ leads to a non-zero relative acceleration of the two geodesics under comparison. This leads to *geodesic deviation*, which has applications in the studies of relativistic tidal forces and the effects of gravitational waves in linearized general relativity [3, 6, 7]

11. The second variation of arclength

Theorem 11.1. *With a metric compatible and torsion-free connection, the second variation of arclength is given by*

$$s''(0) = \varepsilon|\boldsymbol{t}_0|^{-1}\boldsymbol{g}(\boldsymbol{t}_0, \nabla_{\boldsymbol{a}}\boldsymbol{a})\Big|_{\tau_0}^{\tau_1} + \int_{\tau_0}^{\tau_1} -|\boldsymbol{t}_0|^{-3}\left(\frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{a}, \boldsymbol{t}_0)\right)^2 + \varepsilon|\boldsymbol{t}_0|^{-1}\left(|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}|^2 + \boldsymbol{R}(\boldsymbol{t}_0, \boldsymbol{a}, \boldsymbol{a}, \boldsymbol{t}_0)\right) d\tau.$$

Proof. Let us continue from the result of the first variation of arclength, Eq. (9.1). Taking the second partial derivative with respect to the parameter α gives

$$\begin{aligned} s''(\alpha) &= \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{a}\boldsymbol{g}\left(\frac{\boldsymbol{t}_\alpha}{|\boldsymbol{t}_\alpha|}, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}\right) d\tau \\ &= \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{a}\left(|\boldsymbol{t}_\alpha|^{-1}\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a})\right) d\tau. \end{aligned}$$

Since the covariant derivative reduces to a partial derivative when acting on a scalar, and $\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a})$ is a scalar, we have from metric compatibility that

$$\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) = a^i\partial_i\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) = a^i\nabla_i\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) = \boldsymbol{g}(\nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) + \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}).$$

As before, we have that $\nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha = \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}$. Applying this on the first term in the rightmost equality and inserting back to the second variation calculation, we proceed to get

$$s''(\alpha) = \varepsilon \int_{\tau_0}^{\tau_1} \left(\frac{-1}{2}|\boldsymbol{t}_\alpha|^{-3}\varepsilon\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha)\right)\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) + |\boldsymbol{t}_\alpha|^{-1}\left(|\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}|^2 + \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a})\right) d\tau.$$

The first term can be expanded by virtue of metric compatibility, the symmetry of the metric and the torsion-free connection such that

$$\boldsymbol{a}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) = \boldsymbol{g}(\nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha, \boldsymbol{t}_\alpha) + \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha) = 2\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{t}_\alpha) = 2\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}).$$

With this and $\varepsilon^2 = 1$, we get

$$s''(\alpha) = \int_{\tau_0}^{\tau_1} -|\boldsymbol{t}_\alpha|^{-3}\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a})^2 + \varepsilon|\boldsymbol{t}_\alpha|^{-1}\left(|\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}|^2 + \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a})\right) d\tau. \tag{11.1}$$

Furthermore,

$$\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) = \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{a}) - \boldsymbol{g}(\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{t}_\alpha, \boldsymbol{a}).$$

The last term in Eq. (11.1) is in turn simplified followingly by recognizing that it contains a part of the Riemann endomorphism:

$$\begin{aligned}\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{a}) &= \boldsymbol{g}(\boldsymbol{t}_\alpha, \mathcal{R}(\boldsymbol{a}, \boldsymbol{t}_\alpha)\boldsymbol{a}) + \boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{t}_\alpha}\nabla_{\boldsymbol{a}}\boldsymbol{a}) \\ &= \boldsymbol{R}(\boldsymbol{t}_\alpha, \boldsymbol{a}, \boldsymbol{a}, \boldsymbol{t}_\alpha) + \nabla_{\boldsymbol{t}_\alpha}\boldsymbol{g}(\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{a}) - \boldsymbol{g}(\nabla_{\boldsymbol{t}_\alpha}\boldsymbol{t}_\alpha, \nabla_{\boldsymbol{a}}\boldsymbol{a}).\end{aligned}$$

If we evaluate s'' on the critical curve with $\alpha = 0$, all terms containing $\nabla_{\boldsymbol{t}_0}\boldsymbol{t}_0$ vanish by the autoparallelity requirement. This is due to $\boldsymbol{t}_0$ being the tangent vector field of the geodesic solving the earlier first variation problem. Additionally, we recall that $\nabla_{\boldsymbol{t}_0}\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{t}_0\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$ for some vector fields $\boldsymbol{u}, \boldsymbol{v}$. The second variation of arclength results in

$$\begin{aligned}s''(0) &= \int_{\tau_0}^{\tau_1} -|\boldsymbol{t}_0|^{-3}\left(\frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{a}, \boldsymbol{t}_0)\right)^2 + |\boldsymbol{t}_0|^{-1}\varepsilon\left(|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}|^2 + \boldsymbol{R}(\boldsymbol{t}_0, \boldsymbol{a}, \boldsymbol{a}, \boldsymbol{t}_0) + \frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{t}_0, \nabla_{\boldsymbol{a}}\boldsymbol{a})\right) d\tau \\ &= \varepsilon|\boldsymbol{t}_0|^{-1}\boldsymbol{g}(\boldsymbol{t}_0, \nabla_{\boldsymbol{a}}\boldsymbol{a})\Big|_{\tau_0}^{\tau_1} + \int_{\tau_0}^{\tau_1} -|\boldsymbol{t}_0|^{-3}\left(\frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{a}, \boldsymbol{t}_0)\right)^2 + |\boldsymbol{t}_0|^{-1}\varepsilon\left(|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}|^2 + \boldsymbol{R}(\boldsymbol{t}_0, \boldsymbol{a}, \boldsymbol{a}, \boldsymbol{t}_0)\right) d\tau.\end{aligned}$$

□

The above equation is customarily entitled the second variation formula. We may simplify it further by dividing $\boldsymbol{a}$ into parts perpendicular and tangential to $\boldsymbol{t}_0$

$$\boldsymbol{a} = \boldsymbol{a}_\perp + \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{a})\boldsymbol{t}_0$$

and writing the Riemann tensor term in the second variation formula as

$$\boldsymbol{R}(\boldsymbol{t}_0, \boldsymbol{a}, \boldsymbol{a}, \boldsymbol{t}_0) = \boldsymbol{R}(\boldsymbol{t}_0, \boldsymbol{a}_\perp, \boldsymbol{a}_\perp, \boldsymbol{t}_0) + \boldsymbol{R}\Big(\boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0\Big). \tag{11.2}$$

Continuing, the symmmetries of the Riemann tensor and the properties of the Riemann endomorphism give

$$\begin{aligned}\boldsymbol{R}\Big(\boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0\Big) &= -\boldsymbol{R}\Big(\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0\Big)\\ &= -\boldsymbol{g}\Big(\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \mathcal{R}\big(\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\boldsymbol{t}_0\big)\ \boldsymbol{t}_0\Big)\\ &= -\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\ \boldsymbol{R}\Big(\boldsymbol{t}_0,\ \boldsymbol{t}_0,\ \boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0\Big)\\ &= -\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\ \boldsymbol{R}\Big(\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \boldsymbol{t}_0,\ \boldsymbol{t}_0,\ \boldsymbol{t}_0\Big)\\ &= -\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\ \boldsymbol{g}\Big(\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{a})\boldsymbol{t}_0,\ \mathcal{R}(\boldsymbol{t}_0,\boldsymbol{t}_0)\boldsymbol{t}_0\Big).\end{aligned}$$

Now $\mathcal{R}(\boldsymbol{t}_0,\boldsymbol{t}_0)$ is zero, so this and hence also the last term in Eq. (11.2) vanishes. That is,

$$\boldsymbol{R}(\boldsymbol{t}_0,\boldsymbol{a},\boldsymbol{a},\boldsymbol{t}_0) = \boldsymbol{R}(\boldsymbol{t}_0,\boldsymbol{a}_\perp,\boldsymbol{a}_\perp,\boldsymbol{t}_0) = \boldsymbol{R}(\boldsymbol{a}_\perp,\boldsymbol{t}_0,\boldsymbol{t}_0,\boldsymbol{a}_\perp),$$

where we have also used the symmetries of the Riemann tensor. Metric compatibility and the autoparallelity requirement yield

$$\begin{aligned}\nabla_{\boldsymbol{t}_0}\boldsymbol{a} &= \nabla_{\boldsymbol{t}_0}(\boldsymbol{a}_\perp + \boldsymbol{g}(\boldsymbol{a},\boldsymbol{t}_0)\boldsymbol{t}_0)\\ &= \nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp + \boldsymbol{g}(\nabla_{\boldsymbol{t}_0}\boldsymbol{a},\boldsymbol{t}_0)\boldsymbol{t}_0\\ &= \nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp + (\nabla_{\boldsymbol{t}_0}\boldsymbol{a})_\parallel\end{aligned}$$

where we see that we must have $(\nabla_{\boldsymbol{t}_0}\boldsymbol{a})_\perp = \nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp$. Observe that if we split the norm of $\nabla_{\boldsymbol{t}_0}\boldsymbol{a}$ into normal and tangential terms, the latter will bring a signature factor ε for timelike curves as follows.

$$\begin{aligned}|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}|^2 &= |(\nabla_{\boldsymbol{t}_0}\boldsymbol{a})_\perp|^2 + |(\nabla_{\boldsymbol{t}_0}\boldsymbol{a})_\parallel|^2\\ &= |\nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp|^2 + \boldsymbol{g}(\boldsymbol{g}(\nabla_{\boldsymbol{t}_0}\boldsymbol{a},\boldsymbol{t}_0)\boldsymbol{t}_0,\ \boldsymbol{g}(\nabla_{\boldsymbol{t}_0}\boldsymbol{a},\boldsymbol{t}_0)\boldsymbol{t}_0)\\ &= |\nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp|^2 + \boldsymbol{g}(\nabla_{\boldsymbol{t}_0}\boldsymbol{a},\boldsymbol{t}_0)^2\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{t}_0).\end{aligned}$$

The signature factor arises from noticing that $\varepsilon\boldsymbol{g}(\boldsymbol{t}_0,\boldsymbol{t}_0) = |\boldsymbol{t}_0|^2$ and $\varepsilon^2 = 1$. We may now proceed to write

$$|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}|^2 = |\nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp|^2 + \varepsilon\left(\frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{a},\boldsymbol{t}_0)\right)^2|\boldsymbol{t}_0|^2.$$

Taking $\boldsymbol{t}_0$ to be of unit length and inserting the above results, the second variation of arclenght simplifies into

$$s''(0) = \varepsilon\boldsymbol{g}(\boldsymbol{t}_0,\nabla_{\boldsymbol{a}}\boldsymbol{a})\Big|_{\tau_0}^{\tau_1} + \varepsilon\int_{\tau_0}^{\tau_1}|\nabla_{\boldsymbol{t}_0}\boldsymbol{a}_\perp|^2 + \boldsymbol{R}(\boldsymbol{a}_\perp,\boldsymbol{t}_0,\boldsymbol{t}_0,\boldsymbol{a}_\perp)d\tau.$$

We now proceed to define the *index form* $\boldsymbol{I}$ by the above simplified second variation formula such that we shall have $s''(0) = \boldsymbol{I}(\boldsymbol{a}_\perp,\boldsymbol{a}_\perp)$, similar to the convention of [4]. Motivated by the physical idea of identifying the norms and derivatives of vectors tangent and normal to a curve with the velocities and accelerations in the corresponding directions, we generalize $\nabla_{\boldsymbol{a}}\boldsymbol{a}$ by a transverse acceleration vector field.

Definition 11.2. Let $\boldsymbol{t}$ be the tangent vector field and $\boldsymbol{x}$ the transverse acceleration vector field to a curve $\gamma(\tau)$. We define **index forms** as

$$\boldsymbol{I}(\boldsymbol{v},\boldsymbol{w}) = \varepsilon \boldsymbol{g}(\boldsymbol{t},\boldsymbol{x})\Big|_{\tau_0}^{\tau_1} + \varepsilon \int_{\tau_0}^{\tau_1} \boldsymbol{g}(\nabla_{\boldsymbol{t}}\boldsymbol{v}, \nabla_{\boldsymbol{t}}\boldsymbol{w}) + \boldsymbol{R}(\boldsymbol{v},\boldsymbol{t},\boldsymbol{t},\boldsymbol{w})d\tau.$$

Other definitions such as those in [5, 9] may drop the boundary term from the definition of the index form. This can be the case for example if index forms are only used with fixed end-point variations. However, it is often the case in Lorentzian geometry that if the start and end regions are not considered pointlike, they must be treated as Lorentzian *endmanifolds.* Despite the requirement that variations vanish at the endmanifold is often present in classical mechanics, where space is a Riemannian manifold, it is highly non-trivial on Lorentzian manifolds. Hence we decide to keep the boundary term as a part of the index form for the sake of Lorentzian applications.

Index forms have some important properties, which we shall now briefly discuss. For more in-depth and technical discussions on the topic see e.g. [2, 4]. For a starting point, recall that the first variation of arclength resulted in extremal curves. When $s' = 0$, the nature of the extremum as a minimum or a maximum can then be determined from the sign of s''. Should the second variation – or the index form arising from it – change sign along say a minimizing geodesic, it would no longer continue as a minimizing curve. This turns out to be the case at conjugate or focal points, leading to a connection between the index form and Jacobi fields. As an illustrative example, consider first a Riemannian manifold, where the signature factor $\varepsilon = 1$. Now the index form will be positive semidefinite everywhere where the curve is locally minimizing. The occurrence of conjugate points means that we have a Jacobi field $\boldsymbol{v}$ for which $\boldsymbol{I}(\boldsymbol{v},\boldsymbol{v}) = 0$ somewhere along the curve. To see this, take for simplicity a fixed end-point variation such that the boundary term vanishes. Using the product rule and the symmetries of the Riemann tensor, the index form becomes

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{v},\boldsymbol{v}) &= \int_{\tau_0}^{\tau_1} \boldsymbol{g}(D_\tau \boldsymbol{v}, D_\tau \boldsymbol{v}) + \boldsymbol{R}(\boldsymbol{v},\boldsymbol{t},\boldsymbol{t},\boldsymbol{v})d\tau \\
&= \int_{\tau_0}^{\tau_1} \frac{\partial}{\partial \tau}\boldsymbol{g}(\boldsymbol{v}, D_\tau \boldsymbol{v}) - \boldsymbol{g}(\boldsymbol{v}, D_\tau^2 \boldsymbol{v}) - \boldsymbol{R}(\boldsymbol{v},\boldsymbol{t},\boldsymbol{v},\boldsymbol{t})d\tau \\
&= \boldsymbol{g}(\boldsymbol{v}, D_\tau \boldsymbol{v})\Big|_{\tau_0}^{\tau_1} - \int_{\tau_0}^{\tau_1} \boldsymbol{g}\left(\boldsymbol{v}, D_\tau^2 \boldsymbol{v} + \mathcal{R}(\boldsymbol{v},\boldsymbol{t})\boldsymbol{t}\right) d\tau.
\end{aligned}$$

If $\boldsymbol{v}$ is a Jacobi field, the second term vanishes since $D_\tau^2 \boldsymbol{v} + \mathcal{R}(\boldsymbol{v},\boldsymbol{t})\boldsymbol{t} = \boldsymbol{0}$ is precisely the Jacobi equation of Eq. (10.4). Now $\boldsymbol{I}(\boldsymbol{v},\boldsymbol{v}) = 0$ if also the boundary term vanishes. This happens if the Jacobi field is zero at the end-points ergo if the points are conjugate.

There are some generalizations that can be made to the idea of the above example. First, the conjugate (or focal) points need not necessarily be the end-points of the integration interval, but may well lie within it. One may always split the integration into finitely many parts over smaller subintervals such that the conjugate (or focal) points eventually become the end-points of one of the subintervals. With Lorentzian manifolds, it should be noted that the interesting curves to study are often the so-called *cospacelike* geodesics. Their defining property is that the subspaces of all vectors perpendicular to the tangent vector of a cospacelike geodesic are spacelike everywhere along the geodesic. As $\varepsilon = -1$ for Lorentzian manifolds and the index form itself contains an ε in each term, the Lorentzian analogue to the above Riemannian example would involve studying the possible *negative* semidefiniteness of the index form along cospacelike geodesics. Thus to cover both Riemannian and Lorentzian manifolds simultaneously, we shall take that $\varepsilon\boldsymbol{I}$ (acting on two contravariant vectors) will not be positive semidefinite if there are occurrences of conjugate or focal points. Finally, the study of what happens to the boundary term in the index form tends to require some care here, too. It should hence be included when studying variations with endmanifolds instead of some fixed end-points.

12. Myers's theorem

There are several formulations of Myers's theorem, with some tying it more or less together with Bonnet's theorem and adding further content to it. Here we are mainly concerned with how the theorem restricts the distance that can be travelled between two conjugate points along a geodesic on a manifold with Ricci curvature bounded from below. For an extended discussion, see for instance the texts [2, 8] upon which this section is largely based. First we need to define two qualities demanded of the Riemannian manifolds the theorem concerns.

Definition 12.1. A Riemannian manifold is **complete** if it is complete as a metric space. That is, using the distance function given by the Riemannian metric, all its Cauchy sequences converge.

Definition 12.2. A manifold is **connected** if it is not a disjoint union of two nonempty open sets. Any two points on a connected manifold can be connected by curve segments that are piecewise smooth.

Theorem 12.3. *(Myers) Let M be a complete, connected Riemannian n-manifold. If the action of the Ricci tensor on some $\boldsymbol{v} \in TM$ is bounded below by a positive constant H such that*

$$R_{jk}v^j v^k \geq (n-1)H,$$

then every geodesic of length at least $\pi/\sqrt{H}$ has conjugate points.

Proof. Suppose γ is a minimizing unit speed geodesic segment starting at $x \in M$ with a tangent vector field $\boldsymbol{t}$ and $\{\boldsymbol{e}_i\}$ an orthonormal basis parallel transported along γ. Let

$$\boldsymbol{v}_i = \sin\left(\frac{\pi\tau}{L}\right)\boldsymbol{e}_i$$

be a proper normal vector field along γ. Let us study the case of proceeding from $\tau = 0$ to $\tau = L$ along γ. From the definition of index forms with $\epsilon = 1$ for Riemannian manifolds, we then get

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{v}_i,\boldsymbol{v}_i) &= \int_0^L \boldsymbol{g}(D_\tau\boldsymbol{v}_i, D_\tau\boldsymbol{v}_i) + \boldsymbol{R}(\boldsymbol{v}_i,\boldsymbol{t},\boldsymbol{t},\boldsymbol{v}_i)d\tau \\ &= \int_0^L \frac{\partial}{\partial\tau}\boldsymbol{g}(\boldsymbol{v}_i, D_\tau\boldsymbol{v}_i) - \boldsymbol{g}(\boldsymbol{v}_i, D_\tau^2\boldsymbol{v}_i) + \boldsymbol{R}(\boldsymbol{v}_i,\boldsymbol{t},\boldsymbol{t},\boldsymbol{v}_i)d\tau,\end{aligned}$$

where we have fixed the endpoints of variation to get rid of the boundary term arising from the second variation of arclength. The first term in the above equation vanishes, since $\boldsymbol{v}_i$ is a proper normal. Hence we get

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{v}_i,\boldsymbol{v}_i) &= \int_0^L -\boldsymbol{g}(\boldsymbol{v}_i, D_\tau^2\boldsymbol{v}_i) + \boldsymbol{g}(\boldsymbol{v}_i, \mathcal{R}(\boldsymbol{t},\boldsymbol{v}_i)\boldsymbol{t})d\tau \\ &= -\int_0^L \boldsymbol{g}\left(\boldsymbol{v}_i, D_\tau^2\boldsymbol{v}_i - \mathcal{R}(\boldsymbol{t},\boldsymbol{v}_i)\boldsymbol{t}\right)d\tau \qquad (12.1)\end{aligned}$$

Next we compute that

$$D_\tau^2 \boldsymbol{v}_i = -\left(\frac{\pi}{L}\right)^2 \sin\left(\frac{\pi\tau}{L}\right) \boldsymbol{e}_i.$$

Inserting this result into Eq. (12.1), we get

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) &= -\int_0^L \sin\left(\frac{\pi\tau}{L}\right) \boldsymbol{g}\left(\boldsymbol{e}_i, -\left(\frac{\pi}{L}\right)^2 \boldsymbol{e}_i - \mathcal{R}(\boldsymbol{t}, \boldsymbol{e}_i)\boldsymbol{t}\right) d\tau \\
&= \int_0^L \sin\left(\frac{\pi\tau}{L}\right) \left(\left(\frac{\pi}{L}\right)^2 \boldsymbol{g}(\boldsymbol{e}_i, \boldsymbol{e}_i) + \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{t}, \boldsymbol{t}, \boldsymbol{e}_i)\right) d\tau \\
&= \int_0^L \sin\left(\frac{\pi\tau}{L}\right) \left(\left(\frac{\pi}{L}\right)^2 - \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{t}, \boldsymbol{e}_i, \boldsymbol{t})\right) d\tau
\end{aligned}$$

Now recall that the Ricci tensor is obtained by tracing over the first and third indices. Hence

$$\sum_{i=1}^{n} \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{t}, \boldsymbol{e}_i, \boldsymbol{t}) = R_{jk} t^j t^k.$$

Noticing that $R(\boldsymbol{e}_n, \boldsymbol{t}, \boldsymbol{e}_n, \boldsymbol{t}) = R(\boldsymbol{t}, \boldsymbol{t}, \boldsymbol{t}, \boldsymbol{t}) = 0$, we see that it suffices to sum only from $i = 1$ until $n - 1$ to get the Ricci tensor. Therefore

$$\begin{aligned}
\sum_{i=1}^{n-1} \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) &= \int_0^L \sin\left(\frac{\pi\tau}{L}\right) \left(\left(\frac{\pi}{L}\right)^2 (n-1) - \sum_{i=1}^{n-1} \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{t}, \boldsymbol{e}_i, \boldsymbol{t})\right) d\tau \\
&= \int_0^L \sin\left(\frac{\pi\tau}{L}\right) \left(\left(\frac{\pi}{L}\right)^2 (n-1) - R_{jk} t^j t^k)\right) d\tau \\
&\leq \int_0^L \sin\left(\frac{\pi\tau}{L}\right) \left(\left(\frac{\pi}{L}\right)^2 (n-1) - (n-1)H\right) d\tau.
\end{aligned}$$

Thus we see that

$$\sum_{i=1}^{n-1} \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) < 0$$

whenever $L > \pi/\sqrt{H}$. This means that at least one of the index forms in the sum must be negative, forcing x to have a conjugate point on a geodesic along γ at a distance of at most $\pi/\sqrt{H}$ away. □

13. Hawking's theorem

Alongside its overall similarity to Myers's theorem, Hawking's theorem also shares the trait that there are several formulations of the theorem. The main difference is that Myers' theorem is a topic of Riemannian geometry, and Hawking's theorem as the analogous Lorentzian result requires more knowledge of the material discussed in sections 7 and 8. After all, the two theorems are concerned with unequal kinds of manifolds and place distinct restrictions on them. The rise of the study of Lorentzian comparison theorems has also led to their application into proving global theorems such as this one. For instance, see [10] for an alternative formulation of Hawking's theorem and a proof relying on volume comparison. For an introductory discussion on some Lorentzian comparison theorems, see for example [9] and for a couple of more modern developments see [11]. Nevertheless, here we examine the following version of Hawking's theorem, based on the first one of those stated in [4].

Theorem 13.1. *(Hawking) Let M be a Lorentzian n-manifold with a spacelike future Cauchy hypersurface A. Suppose $\gamma(\tau)$ is a timelike curve starting on A with a tangent vector field $\boldsymbol{t}$ and that A has future convergence*

$$k \equiv \boldsymbol{k}^*(\boldsymbol{t}_{\tau=0}) \geq b > 0$$

bounded from below by a positive constant b. If

$$R_{\mu\nu} w^\mu w^\nu \geq 0$$

for every timelike $\boldsymbol{w} \in TM$, then no future timelike curve starting somewhere on A will take a proper time greater than $1/b$.

Proof. Let $y \in D^+(A)\backslash A$. Then there is a timelike geodesic $\gamma(\tau)$ between A and y. As can be done for timelike curves, we identify the length of γ with a proper time interval $\Delta\tau_{A\to y}$. Let $\boldsymbol{t}_\tau$ denote the tangent vector of γ at $\gamma(\tau)$. Suppose γ is defined within the interval $\tau \in]0, 1/\boldsymbol{k}^*(\boldsymbol{t}_0)]$ and has the starting point $x = \gamma(0) \in A$. Denote a local orthonormal basis for T_xM by $\{\boldsymbol{e}_\mu\}$, $\mu \in \{0, ..., n-1\}$, so that $\boldsymbol{e}_0$ is timelike and $\{\boldsymbol{e}_i\}$, $i \in \{1, ..., n-1\}$ are spacelike. Parallel translating the basis along γ results in a new basis $\{\boldsymbol{e}'_\mu\}$ so that $\{\boldsymbol{e}'_i\}$ are again spacelike. We then have the proper variations $\boldsymbol{v}_i \equiv f\boldsymbol{e}'_i$ by scaling the new spacelike basis vectors with a positive function $f(\tau) = 1 - k\tau$, $\tau \in [0, 1/k]$. In this case the transverse acceleration vector field is given by $(\nabla_{\boldsymbol{e}'_i}\boldsymbol{e}'_i)_\perp$, so we have from the definition of the index form that

$$\varepsilon \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = \boldsymbol{g}(\boldsymbol{t}, (\nabla_{\boldsymbol{e}'_i}\boldsymbol{e}'_i)_\perp)\Big|_0^{1/k} + \int_0^{1/k} \boldsymbol{g}(D_\tau \boldsymbol{v}_i, D_\tau \boldsymbol{v}_i) + \boldsymbol{R}(\boldsymbol{v}_i, \boldsymbol{t}, \boldsymbol{t}, \boldsymbol{v}_i) d\tau.$$

Writing the vector $\boldsymbol{v}_i$ as $\boldsymbol{v}_i = f\boldsymbol{e}'_i = (1 - k\tau)\boldsymbol{e}'_i$ and invoking the symmetries of the Riemann tensor, the above equation becomes

$$\varepsilon \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = \boldsymbol{g}(\boldsymbol{t}_0, (\nabla_{\boldsymbol{e}'_i}\boldsymbol{e}'_i)_\perp)\Big|_0^{1/k} + \int_0^{1/k} \boldsymbol{g}(-k\boldsymbol{e}'_i, -k\boldsymbol{e}'_i) - \boldsymbol{R}(f\boldsymbol{e}'_i, \boldsymbol{t}, f\boldsymbol{e}'_i, \boldsymbol{t}) d\tau.$$

Small simplifications and recognizing the second fundamental form $(\nabla_{\boldsymbol{e}'_i}\boldsymbol{e}'_i)_\perp = \boldsymbol{I\!I}(\boldsymbol{e}'_i, \boldsymbol{e}'_i)$ give us

$$\varepsilon \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{I\!I}(\boldsymbol{e}'_i, \boldsymbol{e}'_i))\Big|_0^{1/k} + \int_0^{1/k} k^2 \boldsymbol{g}(\boldsymbol{e}'_i, \boldsymbol{e}'_i) - f^2 \boldsymbol{R}(\boldsymbol{e}'_i, \boldsymbol{t}, \boldsymbol{e}'_i, \boldsymbol{t}) d\tau.$$

Since the upper boundary in the first term vanishes, we have

$$\varepsilon \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = -\boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{I\!I}(\boldsymbol{e}'_i, \boldsymbol{e}'_i)) + k^2\tau \boldsymbol{g}(\boldsymbol{e}'_i, \boldsymbol{e}'_i)\Big|_0^{1/k} - \int_0^{1/k} f^2 \boldsymbol{R}(\boldsymbol{e}'_i, \boldsymbol{t}, \boldsymbol{e}'_i, \boldsymbol{t}) d\tau.$$

Because $\{\boldsymbol{e}'_i\}$ is a set of spacelike vectors, we may sum over $i \in \{1, ..., n-1\}$. Proceeding similarly as in the proof of Myers's theorem, we get

$$\varepsilon \sum_{i=1}^{n-1} \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = -\sum_{i=1}^{n-1} \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{I\!I}(\boldsymbol{e}'_i, \boldsymbol{e}'_i)) + \sum_{i=1}^{n-1} k - \int_0^{1/k} f^2 \sum_{i=1}^{n-1} \boldsymbol{R}(\boldsymbol{e}'_i, \boldsymbol{t}, \boldsymbol{e}'_i, \boldsymbol{t}) d\tau.$$

As $\boldsymbol{R}(\boldsymbol{e}'_0, \boldsymbol{t}, \boldsymbol{e}'_0, \boldsymbol{t}) = 0$, we may extend the sum in the last term. This yields

$$\varepsilon \sum_{i=1}^{n-1} \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = (n-1)k - \int_0^{1/k} f^2 \sum_{\mu=0}^{n-1} \boldsymbol{R}(\boldsymbol{e}'_\mu, \boldsymbol{t}, \boldsymbol{e}'_\mu, \boldsymbol{t}) d\tau - (n-1)\boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{h}_x).$$

Inserting the convergence $k = \boldsymbol{g}(\boldsymbol{t}_0, \boldsymbol{h}_x)$, the first and last terms on the right hand side of the above equation cancel. This leaves us with

$$\varepsilon \sum_{i=1}^{n-1} \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) = -\int_0^{1/k} f^2 R_{\mu\nu} t^\mu t^\nu d\tau.$$

Now since $\boldsymbol{t}$ is timelike, we have $R_{\mu\nu} t^\mu t^\nu \geq 0$. It also holds for a positive f that $f^2 > 0$ for all $\tau \in [0, 1/k]$. Therefore there is some $i \in \{1, ..., n-1\}$ for which $\varepsilon \boldsymbol{I}(\boldsymbol{v}_i, \boldsymbol{v}_i) \leq 0$. Thus some point $\gamma(\tau_1)$ along γ is focal to the starting point in A such that $\tau_1 \in]0, 1/k]$. Notice that $k \geq b > 0$ implies $1/k \leq 1/b$, so the existence of a focal point in the interval $]0, 1/k]$ means that there will certainly be at least one focal point point along γ before y if we let $\tau \geq 1/b$, given that γ can be extended into the interval $]0, 1/b]$ in case $k > b$. Therefore $1/b$ gives an upper limit for the proper time that may elapse in the future Cauchy development of A and we have

$$D^+(A) \subset \{x \in M | \Delta\tau_{A \to y} \leq 1/b\}.$$

Since A is a future Cauchy hypersurface, it holds that $H^+(A) = \emptyset$. This is due to the defining requirement of Cauchy hypersurfaces that there are no timelike inextensible curves that would not pierce the Cauchy hypersurface. A non-empty $H^+(A)$ then again would give rise to the existence of points via which there is no timelike inextensible curve through A. On the other hand, all future timelike curves starting on A belong to its causal future $I^+(A)$. Assume there is a curve $\sigma \in I^+(A)$ defined on some interval $[\tilde{\tau}_1, \tilde{\tau}_2]$ such that there is a $\tilde{\tau} \in [\tilde{\tau}_1, \tilde{\tau}_2]$ for which $\sigma(\tilde{\tau}) \notin D^+(A)$. However, if such a future timelike curve starting in A is to ever escape $D^+(A)$, it should meet the boundary of $D^+(A)$, leading contradictorily to $H^+(A) \neq \emptyset$. Thus $I^+(A) \subset D^+(A)$, so we must have $\Delta\tau_{A \to q} \leq 1/b$ for all future timelike paths starting at A. □

14. Physical implications of Hawking's theorem

Mathematically speaking, the essential result in Hawking's theorem is that the timelike geodesics that pierce some achronal set on a spatially contracting spacetime converge to a single event in finite proper time. Then again, when it comes to physics, the current consensus is that the observed Universe is expanding rather than contracting. Now recall that there is intrinsically no reason to prefer either one of the two possible time orientations of a spacetime. Thus we may reverse the time orientation and yield a statement equivalent to Hawking's theorem, but which is of interest to the cosmological models aiming to describe the observable Universe: timelike geodesics in an expanding spacetime fulfilling the criteria in Hawking's theorem originate from a single event in the past and there is a maximum limit for the proper time measured by any observer on any of these paths. Given that the achronal set covers the whole spatial Universe, the conclusion is that there is an upper limit to the age of the Universe. The idea that an expanding Universe would not necessarily be eternal was around ever since the development of such spacetime models, but Hawking's theorem endows us with some conditions by which some expanding spacetimes can be categorized as originating from a past singularity.

But why did we choose to work in terms of future convergence whilst proving the theorem? This is in fact a fairly standard procedure and a good convention when proving theorems that have to do with the past. One reason lies in the possibility to consider positive time intervals and another one in more psychological or philosophical matters: it is somehow comfortable to consider the forward problem of whether or not a given achronal set will end up in a singularity rather than the inverse problem of whether or not it came from one. However, there is a disclaimer about singularities yet to be stated: singularities are outside the domain of general relativity, and many expect some new physics to arise where classical general relativity would lead to singularities. Nature ought not to break down if our contemporary models and calculations do. There is really no good definition for a physical singularity. Thus we should, in mathematicians' terms, rather refer to geodesic incompleteness when discussing singular behavior.

The requirement that $R_{\mu\nu}v^\mu v^\nu \geq 0$ whenever v^μ are the components of a timelike vector is known as the *timelike convergence condition.* Now proceed by setting the cosmological constant Λ in Einstein's field equations to zero, contract to solve for the Ricci scalar R in terms of the energy-momentum tensor's trace T and insert R back to the field equations to write it as

$$R_{\mu\nu} = \kappa \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right).$$

Next, contracting both sides twice by the components of a timelike vector $\boldsymbol{v}$ and imposing the timelike convergence condition results in an inequality for the energy-momentum tensor known as the *strong energy condition*

$$T_{\mu\nu}v^\mu v^\nu + \frac{1}{2} T v^\mu v_\mu \geq 0.$$

This and other similar inequalities known as *energy conditions* serve as a means of setting limits for the possible physical processes that may happen in spacetime. For instance, the requirement that energy may not flow faster than the speed of light can be set using such conditions. In particular, the strong energy condition displayed here is a way to exclude large negative pressures in some models implying the existence of dark energy and to demand that gravitation is attractive [12, 6]. In general, the non-zero matter and energy content of the observable Universe make it interesting to study the conditions involving the energy-momentum tensor. Nevertheless, the requirements set in Hawking's theorem regarding spacetime curvature seem physically meaningful for the so-called *Robertson-Walker* models used for approximating our Universe. The power of Hawking's theorem is then in removing the need for any certain Robertson-Walker model or its dependency on global assumptions such as spatial isotropy [4].

References

[1] Loring W. Tu. *An Introduction to Manifolds.* Springer, 2011. ISBN: 978-1-4419-7399-3.

[2] John M. Lee. *Riemannian Manifolds: An Introduction to Curvature.* Springer, 1997. ISBN: 978-0-387-98322-6.

[3] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation.* W. H. Freeman, 1973. ISBN: 0716703343.

[4] Barrett O'Neill. *Semi-Riemannian Geometry With Applications to Relativity, 103, Volume 103 (Pure and Applied Mathematics).* Academic Press, 1983. ISBN: 0125267401.

[5] Kai S. Lam. *Fundamental Principles of Classical Mechanics: A Geometrical Perspective.* World Scientific, 2014. ISBN: 978-981-4551-48-9.

[6] Sean M. Carroll. *Spacetime and geometry: An introduction to general relativity.* 2004. ISBN: 0805387323, 9780805387322. URL: http://www.slac.stanford.edu/%20spires/find/books/www?cl=QC6:C37:2004.

[7] M.P. Hobson, G.P. Efstathiou, and A.N. Lasenby. *General Relativity: An Introduction for Physicists.* Cambridge University Press, 2006. ISBN: 9781139447546. URL: https://books.google.fi/books?id=xma1QuTJphYC.

[8] Jeff Ebin David G.; Cheeger. *Comparison Theorems in Riemannian Geometry.* AMS Chelsea Publishing, 1975. ISBN: 978-0-8218-4417-5.

[9] J.K. Beem and P.E. Ehrlich. *Global Lorentzian geometry.* Pure and applied mathematics. M. Dekker, 1981. URL: https://books.google.fi/books?id=4kc1AQAAIAAJ.

[10] Jan-Hendrik Treude and James D. E. Grant. "Volume Comparison for Hypersurfaces in Lorentzian Manifolds and Singularity Theorems". In: *Annals Global Anal. Geom.* 43.3 (2013), pp. 233–251. DOI: 10.1007/s10455-012-9343-z. arXiv: 1201.4249 [gr-qc].

[11] Debora Impera. "Comparison theorems in Lorentzian geometry and applications to spacelike hypersurfaces". In: *Journal of Geometry and Physics* 62.2 (2012), pp. 412–426. ISSN: 0393-0440. DOI: https://doi.org/10.1016/j.geomphys.2011.11.004. URL: http://www.sciencedirect.com/science/article/pii/S039304401100252X.

[12] S.W. Hawking and G.F.R. Ellis. *The Large Scale Structure of Space-Time.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1973. ISBN: 9780521099066. URL: https://books.google.fi/books?id=QagG%5C_KI7Ll8C.